



# Socio-economic status and household catastrophic health expenditure in urban Argentina

*Departamento de Economía*

*Maestría En Econometría*

Mercedes Mac Mullen

*Junio de 2015*

## Abstract

**Objectives:** is to explore the socioeconomic determinants of catastrophic out-of-pocket health expenditure in Argentina in order to be able to capture the distribution of catastrophic health expenditure (CHE) among health coverage.

**Methods:** A cross-sectional study was conducted on the Household Expenditure Survey for 2012/2013. Using the household's capacity to pay (CTP) the proportion of households with CHE was estimated as those households that face health out-of-pocket above 40% of their CTP. A logistic regression was used to identify the determinants of CHE. To measure socioeconomic status a composite wealth index was calculated.

**Results:** The final sample included 20,699 households. The main determinants were having a senior member [OR=2.14 (1.40-3.28)], senior head of house hold [OR=0.46 (0.29-0.74)] and employment status [OR=0.63 (0.49-0.80)]. Stratified analysis by health coverage was not statistically significant.

**Conclusions:** There are few households that had health OOP above 40% of their capacity to pay, over 97% of the households had OOP that were below this threshold. Moreover, only 51% of the households incurred in out-of-pocket health payments. Impoverishment rate after OOP payments was 0.8%, higher for the lower quintiles.

## Contents

Abstract.....	2
1. Introduction .....	5
2. Background .....	6
2.1. Argentinean Health Care System .....	6
2.2. Other health care expenditures and health system composition information .....	8
2.3. Literature Review .....	9
3. Objectives and Hypothesis.....	11
3.1. Main and secondary objectives .....	11
3.2. Hypothesis.....	11
4. Materials .....	12
4.1. Study design and data source .....	12
4.2. Survey reliving and design .....	13
4.3. Observation Units .....	13
4.4. Expenditure Data .....	14
4.5. Study population.....	16
5. Concepts and definitions .....	16
5.1. <b>Dependent variable: catastrophic out-of-pocket household health expenditures (CHE)</b> .....	16
5.2. <b>Independent variables</b> .....	17
5.2.1. <b>Socioeconomic Status</b> .....	18
5.3. <b>Catastrophic Health Expenditure</b> .....	19
6. Wealth Index.....	21
6.1. Methods.....	21
6.2. Statistical Analysis.....	23
6.3. Results.....	24
7. Catastrophic health expenditure logistic model.....	27
7.1. The Logistic regression model <sup>39,40</sup> .....	27
7.2. The logistic regression applied to catastrophic health expenditure1.....	33
7.3. Statistical analysis .....	35
7.4. Results.....	35
7.4.1. <b>Descriptive analysis</b> .....	35
7.4.2. <b>Determinants of Catastrophic Health Expenditure</b> .....	36

**7.4.3. Logistic regression analysis.....38**

8. Discussion and conclusions..... 40

9. References ..... 44

10. Tables ..... 47

11. Figures..... 53

## 1. Introduction

Argentina has a fairly well developed health system considering standards in developing countries.<sup>1</sup> However, a number of Argentina's health status indicators are worse than those of middle income countries in the region with lower health expenditure and per capita income, such as Chile, Costa Rica and Uruguay.<sup>1,2</sup>

A comparative study from the Economic Commission for Latin America and the Caribbean (Petricara, M. 2008) shows that out-of-pocket health expenditure in the Argentina is one of the largest in the region. In Brazil, Chile, Mexico and Uruguay is approximately 5% of capacity to pay. For Argentina and Ecuador, is substantially greater, 8% and 7.4% respectively.<sup>2,3</sup>

Furthermore, Argentina is the country that has one higher percentage of households that incur catastrophic expenses on health.<sup>2</sup> However, at the same time Argentina is the country which reports a lowest proportion of households with out-of-pocket health expenses in the region. Overall, 50% of households do not report such expenses.<sup>2</sup> In Chile, 37% of households have zero pocket expenditure; in Brazil, Colombia, Mexico and Uruguay this ratio lies between 20 and 30%.<sup>2</sup>

Argentina is clearly a special case, where the families make in average high OOP payments, but more than half of them reported no out-of-pocket expenses.<sup>2</sup> This means that in Argentina, those who have positive health out-of-pocket expenditure, spend on average 16% of their ability to pay for this concept.<sup>2</sup>

Various reasons can make a home have out-of-pocket expenses in health:<sup>2</sup>

- (i) a good coverage of health (through public or private systems);
- (ii) do not experience events of morbidity during the measurement period;
- (iii) that the household has a precarious income level, and does not have access to credit, therefore be prevented from spending on health.

If a home is restricted in their spending (reason iii), there is the real value of their need for expenditure on health, but in this case health payments are censored. And the origin of the censorship in this variable is not random, but potentially depends on the same set of determinants of OOP payments.<sup>2</sup>

Comparing the percentage of households that are "impoverished" after dealing with health costs across the region, Uruguay, Mexico and Brazil show very low rates (below 0.6%), whereas in Argentina this rate particularly high, 1.9%. For the rest of the countries it is close to 1%.<sup>2,3</sup>

In Argentina, those households with complete health coverage are the ones that have a higher burden of health expenditure and suffer of higher impoverishment rates.<sup>1,2</sup> Those households with health insurance destine one third of their monthly total expense to health payments.<sup>2</sup>

An overview of the Argentine health system and literature review will be presented in the next section. In Section 3 the objectives and hypothesis of the study will be presented. Section 4 provides information on the study design and data source, including the sampling methods for the National Household Expenditure Survey. Section 5 describes the variables included in the design and how they are operationalized. Section 6 elaborates a detailed description of the methods, analysis and results for the estimation of the Wealth Index. In section 7 the logistic regression model is described and applied for our study, also the results for the study are presented in this section. Section 8 discusses the empirical findings of the study.

## 2. Background

### 2.1. Argentinean Health Care System

In 2005 Argentina spent 10.2% of its Gross National Product on health care<sup>4</sup>. According to WHO's World Health Report 2000, Argentina fared poorly considering their health and human resources, ranking 36th in health level after Cuba, Chile and Uruguay, 89th in equity in financing and 75th in overall health system performance, behind many countries in the Region<sup>5</sup>. The current health care system has undergone a long history of change and modifications. These modifications were conducted in the context of major state reforms that carried with them a growing increase in external debt, a gradual withdrawal of the state as the generator of social services, an unequal distribution of income, economic growth and the exchange of goods and services directed by the open market. The last major health system reform in the 1990 intended to insight a competitive market between the social health insurance system and the private sector. In addition to the competition between the varying funds and benefit packages offered by social health insurance and private health care plans, it aimed to decentralize hospital management, and address the mounting deficit of the pensioners' health plan (PAMI)<sup>1</sup>. At present the Argentine health care system is comprised of three sub-systems: (1) the

public health care system, (2) the social health insurance system or obras sociales or union plans and (3) the private health care sector.

**Public health care system.** The public health care system in theory grants free access to health care services to every citizen in the country (universal health care system). However, the services are used mainly by the population with fewer economic resources. The public system is financed through public taxes and after the 1990's health care system reform people can be asked for a fee for service.<sup>1</sup> According to the 2001 population census, 48.1% of the population is without medical coverage from a social health insurance plan and/or private insurance.<sup>6</sup> The public system designates 30% of their expenditure to preventative programs and 70% to medical attention. Through the reform the publicly funded sector was highly decentralized from a federal level to the provincial or municipal level where the ministers of health from each province are now in charge of the provision of basic health services, including prevention, education and promotion of the health. In addition the restructuring granted public hospitals the ability to provide service to uninsured as well as provide fee for service to those with private and social insurance.<sup>1</sup> The introduction of the co-payment for the provision of basic health care services, represents an out-of-pocket expenditure assumed by the patient.

**The social insurance system (*Obras Sociales*)** Is made up of approximately 300 plans formed by trade unions or workers unions associated with different industrial sectors classified or organized as national or syndicate, provincial, a retirement health plan (PAMI), independent, and individual companies. This system is tied to the changes of the labor market and is financed by a compulsory contribution made by the employee (3%) and the employer (6%). It was estimated in 2001 that almost 46.7% of the population was covered by a social insurance plan. However, only 20 organizations (7%) covered 40% of the beneficiaries. The social insurance plans are associated with their industrial sectors. This resulted in contrasts between the plans according to the number of workers belonging to the group or industrial sector and the average wages earned that would contribute to the fund, as such the quality and amount of the administered services (rich industrial sectors versus poor sectors) varied significantly. In order to address the issue of inequities among the insurance funds and obtain a partial compensation that would give workers similar benefit packages, a Redistribution Fund was created (Fondo Solidario de Redistribución, FSR). Only workers in the formal work sector are covered by this system. As most social insurance plans are too small and do not have the appropriate infrastructure to attend to their beneficiaries, they end up subcontracting services from the private and public system.

**PAMI.** The PAMI social insurance plan is a special case and is distinct from the other plans. It is a separate health insurance fund for retired individuals. Services are financed through compulsory contribution of workers' wages and contributions of pension benefits (3% from compulsory retirement governmental plan, 1% from workers salary and 1.5% contributed by the employer).

**Private health sector.** The private health care sector is 70% financed through organized prepaid medical plans, 15% by direct and voluntary pre-payments by the social insurance plans and 15% by mutual benefit companies and community hospitals. The health care expenditure of the private sector represents 44% of the total health care spending, of which 80% goes to direct costs of medications and supplies and 20% to private insurances premiums. The private insurance is estimated to cover about 10% of the population or nearly 4 million people.<sup>1</sup>

## 2.2. Other health care expenditures and health system composition information

Of the over 10% of the GDP allocated to health care, the public sector represents 1.54%, the social insurance sector 2.92% and the private sector 7.71%. Between 1995 to 2001 the composition of the cost in health was 43% for the private sector, 34% for the social system and 22% to the public sector. In addition, the human resource pool of the health care system has approximately 110,000 doctors (of which 55% practice in CABA and the Province of Buenos Aires), 42,000 nurses, 16,000 establishments (approximately 7000 belonging to the public sector), around 160,000 beds in the three subsystems, 457 x-ray machines (50% in CABA and Greater Buenos Aires) and 115 ultrasounds (65% in CABA and Greater Buenos Aires).

One study demonstrated that 65.1% had health care coverage aside from the public health care access.<sup>7</sup> The study showed a higher association between health care coverage and hospital admissions, higher education and older age, while findings indicated no difference in coverage by sex.<sup>7</sup> At a national level the study revealed that 23.5% of people with a health care problem did not seek treatment due to lack of access to services, 11.7% in the province of Buenos Aires where there is higher stated health care coverage with private and social insurance.<sup>7</sup> Of the consults conducted in the hospitals 55.6% correspond to the residents from within city, while 41.2% correspond to residents of the province of Buenos Aires and 3.2% are from outside the province.



### 2.3. Literature Review

Health public policy tends to be aimed not only at improving the health status of the population but also to ensure that health expenditures do not have an unsustainable burden on families. Berki (1986) states that health spending becomes financially at risk turning catastrophic when it threatens the ability of families to maintain their standard of living.<sup>8</sup> More precisely, the World Health Organization (WHO) estimates that families that face health expenditures that represent a large proportion of its expenditure after food purchases have a great possibility of being impoverished in time.

Cavagnero et al. (2006) analyses determinants of health service utilization and probability of incurring catastrophic OOP payments in Argentina. This study analyzes household level consumption expenditure data collected by the National Household Expenditure Survey for year 1996-1997. The paper brings out the importance of medical insurance in reducing the probability of catastrophic OOP health payments. The results also show that the presence of senior members, education of head of the household and higher income increase the probability of incurring catastrophic OOP health expenditure. In this study, catastrophic OOP health expenditure is said to occur if OOP health expenditure exceeds 40 percent of capacity to pay (income after food consumption).<sup>1,9</sup>

Knaul et al. (2011) Compares patterns of catastrophic health expenditures in 12 countries in Latin America and the Caribbean using most appropriate available household survey for each country. To measure out-of-pocket health expense two types of prevalence indicators were used: a) relative to an international poverty line, and b) relative to the household's ability to pay net of their food basket. Ratios of catastrophic expenditures were estimated across subgroups defined by economic and social variables. The results show that the percent of households with catastrophic health expenditures ranged from 1 to 25% in the twelve countries. In general, rural residence, lowest quintile of income, presence of older adults, and lack of health insurance in the household are associated with higher propensity of catastrophic health expenditures. However, there is vast heterogeneity by country.<sup>10</sup>

Results for this study find that Argentina was amongst the countries with the highest proportion of catastrophic health expenditure, 8.4%. Argentina and Dominican Republic are the only two countries in the region where large households have lower prevalence of catastrophic health expenditures than small households.

Perticara (2008) characterized OOP health expenditure incidence in Argentina, Brasil, Chile, Colombia, Ecuador, México and Uruguay. For the case of Argentina the NHES 1996-1997 is analyzed.

This study defines catastrophic health expenditure following to Xu et al. (2003). Findings were in line with those from Cavagnero and Knaul. Argentina shows the highest incidence of catastrophic health expenditure. The author highlights the strong association between CHE and having an elderly member in the household.

In general, more developed countries are protected from catastrophic costs through the health system. However, developing countries often have systems that generate large direct costs to families, since they do not provide mechanisms for solidarity, and therefore do not involve third parties (insurance or state) and are charged directly to that provided healthcare cost.<sup>11</sup> This situation leads to the existence of catastrophic health expenditure, being higher in low-income families where there is a disabled, retired family, or have a chronic illness.<sup>12</sup>

In addition, the literature uses the catastrophic health expenditure (CHE) as a measure of system performance. It is proved that those who have some type of health insurance reduce the probability of incurring in risky OOP.<sup>11</sup> Knaul et al. (2011) find that this does not hold in Argentina. For the majority of the countries in Latin America and the Caribbean, the propensity to incur in catastrophic spending is higher among households without insurance. The exceptions to this are Argentina and Chile, Costa Rica and Peru, where those without insurance have lower chance on incurring in catastrophic health expenditure.<sup>10</sup>

There is evidence in the literature that suggests that welfare is reduced by uncertainty in health spending. Families may be forced to borrow to cover unexpected medical expenses with the risk of being trapped in a long-term debt.<sup>13,14</sup> Consequently out-of-pocket (OOP) spending can have great impact on the standard of living of families. In many cases families may have too restrictive budgetary constraints forcing them to sacrifice spending on necessary goods, leading to even more poverty.<sup>13-15</sup> In other cases families will be discourage to use health services, leading to a negative impact on health outcomes for these families.<sup>8,16-22</sup>

Wagstaff and Doorslaer (2003) propose and implement a variety of measures that capture the incidence and intensity of catastrophe in health spending Vietnam in 1993 and 1998. They find find that the poverty impact of out-of-pocket payments is primarily due to poor people becoming even poorer rather than the non-poor being made poor, and that it was not expenses associated with inpatient care that increased poverty but rather non-hospital expenditures.<sup>15</sup> This result was supported by other studies in several countries.<sup>13,20,21,23</sup>

The definition of CHE varies depending on how this variable is operationalized. Wagstaff and Doorslaer (2003) defined catastrophic health expenditure as a share of OOP payment on health of more than 10% of total consumption, including expenditures on both food and non-food items.<sup>15</sup> Xu et al. (2003) defines OOP healthcare expenditure as ‘catastrophic’ if it exceeded 40% of household non-food expenditure or capacity to pay in the past 30 days.<sup>11,24</sup> For the present study the later definition will be used.

### 3. Objectives and Hypothesis

#### 3.1. Main and secondary objectives

Although Argentina has universal health access, families may still incur in significant health expenses. A study on the determinants of catastrophic health expenditure with 1997/98 data showed that 5.5% of the families experience CHE. This disproportioned OOP led 1.7% households into poverty.<sup>1</sup>

1. The main purpose of the present study is to explore the socioeconomic determinants of catastrophic out-of-pocket health expenditure in Argentina 2012-2013 in order to be able to capture the distribution of catastrophic health expenditure among different socioeconomic groups.
2. Secondly, the study seeks to understand why some households still experience catastrophe and impoverishment due to OOP payments under universal health coverage. Hence, we will explore if there are any differences in the distribution of CHE among type of coverage.
3. Measure socioeconomic status in Argentina. The main objective entailed building an asset and housing based socio-economic status index (we will call it wealth index). This index was used to determine the relationship between household socio-economic characteristics and inequalities of catastrophic health expenditure.

#### 3.2. Hypothesis

Based on this previous study from Cavagnero et. al. (2006) and the fact that poor families may be discouraged to pay for health services if they represent a large proportion of their capacity to pay (they might have to sacrifice food spending in order to pay for health care), the hypothesis is that

households in the second, third and fourth quintile will face higher rates of CHE, than other socioeconomic groups.<sup>9,16</sup>

## 4. Materials

### 4.1. Study design and data source

A cross-sectional analysis will be carried out based on the National Household Expenditure Survey 2012/2013 (NHES2012). The survey was conducted across the country between the months of March 2012 and March 2013, through a probabilistic, multistage, stratified sample from Master Urban Sample of Households of Argentina. From this master sample, constructed from the 2010 National Census of Population and Households, around 37,000 homes were selected.

The survey was conducted under the principle of methodological centralization at the National Bureau of Statistics (Instituto Nacional de Estadísticas y Censo – INDEC) and executive decentralization in the Provincial Bureaus of Statistics. Each province was responsible for conducting the survey, coding and data entry, and the first instance of quality control. The INDEC was in charge of methodological and conceptual design and, in its capacity as coordinator of the National Statistical System, carried out the consistency, purification and final editing of the data.

This survey provides information on the social and economic situation of urban population of the all the provinces, as well as on the composition of household spending. The aim of the survey is to provide information on the living conditions of the general population and household groups in particular from the point of view of their participation in the distribution and procurement of goods and services.

The survey has four specific objectives:

- 1) To estimate the structure of household expenditure, source of income and its allocation to the various expenses
- 2) Provide information for assessing the living standards of households and to facilitate programming social spending.
- 3) Provide basic information for estimates of national accounts, in particular concerning the accounts of the household sector and the development of the Supply and Use Tables.

4) Provide information necessary to construct the weighting structure of a Price Index National urban consumer.

These data were collected by administering 5 questionnaires: Household Characteristics; daily expenses; miscellaneous expenses; personal expenses; and Income

#### 4.2. Survey reliving and design

Each home was under study for a week, from Sunday to Saturday. During this week, they were visited in three opportunities to conduct the interviews.

Expenditure information was obtained using a combination of two methods of recruitment. For regular expenses (power, public transportation, cigarettes, etc.), during the week of the survey member of the household were requested to place annotations on the questionnaires. In the case of other expenses, interviews were conducted in which the homes were asked to recall the expenses incurred during different periods of reference (in the last month, two months, six months or in the last year, according to the type of expenditure).

Information pertaining to income was relieved for a reference period of six months. This was obtained through an interview to each income recipient or to a qualified member during the last visit to the home.

Finally, the information corresponding to the socio-demographic characteristics and residential household was obtained through the administration of a questionnaire for the household and its members in the opening interview.

#### 4.3. Observation Units

Observation units of the survey are *private households* living in private homes located in towns of more than five thousand inhabitants. Private household is considered to that made by any person or group of people, relatives or not, who live in the same house under a regime of family type and eat food with the same budget (3).

People who live in the same house under a regime of family type, share their food expenses and living in the household for 6 or more months or, if they lived in it less than six months but have been fixed or plan to set up residence there, are considered *household members*.

The main study variables of the survey are spending and household income. To define and analyze and characterize different domains and characterize households that comprise them, information on demographic, occupational and educational of its members are also gathered, as well as the characteristics of housing, transfers in kind received and household equipment.

#### 4.4. Expenditure Data

The NHES2012 relived information on current expenses, net accumulation of assets and other uses of household resources, excluding expenses related to their business.

The criterion for recording of expenditures is the acquired expenditure. Therefore, the survey registers the value of the goods and services on which the home takes possession (or acquire the right to receive) during the reference period, regardless of the time the household cancels the purchase and the period during which they are consumed.

*Current expenditure* comprises the final consumption expenditure of households and non-consumption expenditures. Consumption spending is the market value of all purchases of goods and services, whether they are paid in cash or credit, by households to meet their needs. It includes procurement of goods and services carried out by members of the household for individual and collective consumption and for gift giving; the goods and services that the home removed from any business or enterprise in its property for its use; the goods and services that any of its members received as payment in kind for their work; and the primary goods produced by the home for their own consumption.

Consumption spending is classified into nine divisions, according to their purpose:

- 1) Food and beverage: includes all food and beverages (alcoholic and non-alcoholic) acquired for consumption inside and outside the home.
- (2) Goods and various services: includes cigarettes, toiletries, personal care services and other goods and services.
- (3) Education: includes educational services for formal education (quota and tariffs for preschool, primary, secondary and University) and non-formal (languages, among others), as well as texts and school supplies.

(4) Equipment and home maintenance: includes goods to furnish the home (furniture, home appliances, household appliances; dishes, white and linen); items for the home maintenance (household cleaning items, tools) and services for the maintenance of the household (furniture repair, repair of appliances and electrical appliances, laundry, laundry, domestic service).

(5) Entertainment: includes tourism services, audio, TV, video and computer equipment, services, leisure activities (sports, cinema, theatre, concerts, quotas of sports club, TV cable, etc.), books, journals and non-professional magazines; and other goods (computers of cinema and photography and accessories, movies, toys and games, pets, sporting goods).

(6) Clothing and footwear: is comprised of clothing (coats, indoor and outdoor clothing for men, women and children), shoes (for men women and children), fabrics, accessories and services (yarn, haberdashery items, repair of clothes and shoes).

(7) Properties, fuel, water and electricity: includes rental housing, common costs and repairs, as well as fuels for use in the home (bottled gas, red gas, firewood, among others), water and electricity.

(8) Health: includes therapeutic accessories and medicinal products (medicines, items for first aid, accessories and equipment) and services (prepaid medical assistance, query me - system health (you medical and dental, internments, childbirth, physiotherapy, clinical and radiological analysis).

(9) Transport and communications: includes the buying and selling of cars, its operation and maintenance (fuel, insurance, parking, etc.), public transport, mail and phone.

The outlays called non-consumption expenditure correspond to transferences the households make without any consideration of any kind for their benefit, as gifts or aid money to other homes within or outside the country, donations in money to not-for-profit institutions, cooperating school, taxes and fines of buildings and motor vehicles, taxes and fines on contracts, legalization, income, personal property, loss of money on account of loss or theft, etc.

The net accumulation of assets includes net purchase of real estate, machinery and equipment for the economic activity of the home, part of inheritance to non-members of the home, jewelry, gold and works of art, and so on.

Other uses of resources are purchases of bonds and public and private bonds, stock purchases and holdings in companies, purchases of foreign currency loans made to non-members of the household, money deposited in warranty, payments of assessments, advances and debts.

#### 4.5. Study population

Data is collected at the household level therefore the unit of analysis will be urban Argentinean households from every province in the country. A total of 20,960 households were included in the study. Individual level variables that are included in the model correspond to the head of the household.

## 5. Concepts and definitions

### 5.1. Dependent variable: catastrophic out-of-pocket household health expenditures (CHE)

The main outcome of this study is catastrophic out-of-pocket household health expenditures (CHE). The household expenditure survey reports expenditures corresponding to the last 30 days prior to the interview. The outcome will be assessed following Xu et al. (2003) methodology<sup>24</sup> where health expenditure is considered catastrophic if it exceeds 40% of the household's capacity to pay.<sup>11,17</sup> To follow this methodology it is necessary to define a few intermediate variables:

i. **Out-of-pocket health expenditure (oop):** refers to all the health expenses the household faces when receiving a health service. The NHES2011 classifies health expenditure into two main broad categories:

- a. Medical products, devices and equipment: this category includes pharmaceutical products, first aid and other medical products, devices, equipment and their repair.
- b. Health services: health services for outpatients, which include medical consultations, dental services, auxiliary services, surgery, childbirth and hospitalizations.

ii. **Total household expenditure (exp):** refers to the monetary value of the total expenditure the household incurs in goods and services. The NHES2011 includes all current expenses, net accumulation of assets and other uses of household resources, excluding expenses related to their business for the month previous to the interview.

iii. **Food expenses (food):** Includes all meals and drinks acquired for consumption within the home. Alcoholic drinks are excluded.

iv. **Household subsistence spending (se):** Refers to the minimum expenditure level a household must incur to maintain a basic lifestyle. This level is determined by a poverty line (defined as



the food expenditure of the household whose food expenditure share of total household expenditure is at the 50th percentile).

**v. Capacity to pay (ctp):** defined as the non-subsistence effective income of the household.

Given the characteristics of the Argentine health system, already detailed in section 2.1, we will use household expenditure data, which is more reliable.<sup>25</sup> Hence, the capacity to pay (ctp) of a household will be determined as follows:

$$ctp_i = \begin{cases} exp_i - se_i & \text{if } se_i \leq food_i \\ exp_i - food_i & \text{if } se_i > food_i \end{cases} \quad (1)$$

Where  $se_h$  accounts for subsistence household expenditure. Poor families tend to have food expenditures below their subsistence level, in some cases this is explained by the existence of food subsidies or because their income level is low enough to not being able to reach minimum subsistence levels of food consumption. To include this in our analysis, we estimated a  $se_h$  for each household adjusting for the household size. For those households below the poverty line (effective food expenditure below the subsistence spending) the CTP was estimated by this threshold, Equation 2.

**vi. Catastrophic Health Expenditure:** Next the out-of-pocket health expenditure was estimated relative to the household's CTP. If this percentage exceeds 40%, then health expenditure is considered catastrophic, and a variable  $che_h$  is defined equal to one:

$$che_h = \begin{cases} 1 & \text{if } \frac{oop_h}{ctp_h} \geq 0.4 \\ 0 & \text{if } \frac{oop_h}{ctp_h} < 0.4 \end{cases} \quad (2)$$

## 5.2. Independent variables

The variables of interest, i.e. determinants of catastrophic health expenditure, included in our model are:

- Head of household's age: Age in years of the head of household. Continuous variable.
- Sex: Sex of the head of household. Binary variable (male/female),
- Education: Highest instruction level achieved by the head of the household categorized as no instruction, primary, secondary, superior, college, other. Categorical variable.

- Employment: occupational situation of the head of household, categorized as unemployed, working XXX. Categorical variable.
- Health coverage: Type of head of household's health coverage, categorized as no coverage, social insurance system, private. Categorical variable.
- Child: Indicates if there are any children under five years in the household.
- Senior: If there are senior members (>65 year old) in the household.
- Senior head of household: If the head of household is over 65 years old.
- Socioeconomic Status: Measured using a Household Wealth Index (the construction of the index will be detailed in the Section XX).

### 5.2.1. Socioeconomic Status

In order to address a measurement of socioeconomic status the literature proposes several proxy measures like consumption, income or asset indicators. It is broadly discussed the reliability of income measurements since often families with low socioeconomic status tend to over-report their income, while families that are better off have a tendency to underreport them.<sup>26</sup> An additional disadvantage about using income as a measurement for SES is that income for self-employed people and agriculture workers have high seasonality and therefore are difficult to account for.<sup>27</sup> Even more, many times income fluctuation in the household (for example negative impact shock due to illness) does not necessarily imply a change in living standards, since they might be able to maintain consumption through savings. Therefore it would be misleading to measure socioeconomic status through income data.<sup>25</sup>

Finally, in countries where the informal labor market is large, expressing income or expenditure levels in monetary values can be extremely time-consuming and suffers important reliability problems.<sup>28</sup> This is the case of Argentina's labor market, where in 2012 informal employment accounted for 34% of the total employed population.<sup>29</sup>

Given economic theory, many economists prefer to measure living standards proxied by consumption.<sup>25</sup> Using consumption to measure wealth inequality possesses its own disadvantages. Usually surveys that relive information on consumption patterns ask about consumptions during the past weeks, months or even year (to avoid seasonality)<sup>30</sup>, this may lead to recall bias mismeasurement.<sup>27</sup>

Other measures of living standard widely used in the literature are wealth or asset index. First of all, information on asset possession and housing conditions does not face recall problems. Further,

many of the items covered in the asset questionnaire, such as houses' construction material and household members physical possessions, can be independently verified through investigators' independent observation, thereby increasing reliability.<sup>30</sup> Secondly, as well as using consumption as a proxy, wealth indexes have a longer term perspective.<sup>25,28</sup> The problem with consumption data is that it is much more expensive to collect compared to asset data.<sup>25</sup> Moreover the typical asset questions can be asked and answered in less than five minutes, which can be added much more easily.<sup>30</sup>

Therefore, in order to measure household socioeconomic status (SES) a Household Wealth Index (HWI) was estimated following the methodology used to estimate the DHS Wealth Index proposed by Filmer et al. (2001) and Rutstein et al. (2004).<sup>31,32</sup> Other advantage of using the Wealth Index to measure SES is that reflects a more permanent status of the socioeconomic position of the household.<sup>33</sup>

McKenzie (2003) acknowledges that asset index has two disadvantages that arise when there are not enough indicator variables to detect wealth inequality among the households.<sup>27</sup> The first one is clumping, which occurs when households are clumped together in a small number of groups. To exemplify this imagine the most extreme case where only one asset indicator is used. In this case there will be only two wealth categories, those who own the asset and those who do not. The second potential problem is truncation of the asset distribution. Truncation may occur when there are not enough asset indicators to differentiate between the poor and the very poor or the rich and the very rich. To detect if the estimated approximation has any of this two potential issues he suggests to analyze the distribution of the index.<sup>27</sup>

The Argentinean Household Expenditure Survey collects plenty of information on living standards and asset position for each household that can be used to measure the economic status. The first step in the construction of the Wealth Index was to identify the indicator variables that will be included and dichotomize them (Table 1).<sup>31,32</sup>

Once all of the indicator variables were recoded, then the second step was to obtain the weights to construct the Index. Principal component analysis (PCA) was carried out to assign the weights for each indicator using the factor loadings for the first component. The whole procedure will be further detailed in the Section 6.<sup>25,27,31,32</sup>

### 5.3. Catastrophic Health Expenditure

In this section the methodology to construct the dependent variable will be detailed following Xu (2005).<sup>24</sup>

Given that the subsistence expenditure depends on the size of the household the first step was to estimate a household size adjustment factor <sup>11,24</sup> by calculating the coefficient of a linear regression for the logarithm of food expenditure explained by the logarithm of the household size:

$$\ln(\text{food}_i) = \ln(k) + \beta \ln(\text{hsize}_i) \quad i = 1, 2, \dots, n \quad (3)$$

Where  $k$  is a constant and  $\text{hsize}$  corresponds to the household size measured by the number of members in the household. For our data the value of  $\beta$  is 0.53 (95%CI 0.516-0.549). This result implies that food consumption increases as the household size increases, but this increase is less than proportionate. This result is similar to the one on Xu et al. (2005) found for an average of 59 countries ( $\beta=0.56$ ). Then we used this coefficient to adjust the household size and obtain an equivalence scale:

$$\text{eqsize}_i = \text{hsize}_i^\beta \quad i = 1, 2, \dots, n \quad (4)$$

Secondly the equivalised per capita food expenditure was estimated by dividing the household food expenditure by the adjustment factor:

$$\text{eqfood}_i = \frac{\text{food}_i}{\text{eqsize}_i} \quad (5)$$

Thirdly, to estimate the subsistence expenditure it was necessary to define the poverty line. In order to accomplish this, the proportion of food expenditure relative to the total household expenditure was estimated for the 45 and 55 percentile.

$$\text{foodexp}_i = \frac{\text{food}_i}{\text{exp}_i} \quad (6)$$

The 45 and 55 percentiles for our sample are 0.331 and 0.380 respectively. Then all the households with a proportion of food expenditure ratio between this range were identified and an average of the adjusted per capita food expenditure was estimated. This average will constitute the poverty line (*povline*). Based on the poverty line and the adjusted household size it was possible to estimate the subsistence household expenditure (*se*):

$$\text{se}_i = \text{povline} * \text{eqsize}_i \quad (7)$$

The next step was to estimate the capacity to pay (*ctp*) using equation 1. Once the capacity to pay was estimated it was possible to obtain the out-of-pocket health payments (*oop*) as a share of household *ctp*:

$$oopctp_i = \frac{oop_i}{ctp_i} \quad (8)$$

Finally it was possible to obtain the catastrophic health expenditure using equation 2.

To determine how many people were pushed into poverty we define<sup>24</sup>

$$impoor = \begin{cases} 1 & \text{if } exp_i \geq se_i \text{ and } exp_i - oop_i < se_i \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

## 6. Wealth Index

### 6.1. Methods

As mentioned above, socioeconomic status of the households will be assessed by a wealth index. The index will be computed using Principal Component Analysis (PCA) following the procedure that it is used for the construction of the wealth index that is included in the Demographic Health Survey (DHS) and the methodology proposed by Filmer et. al (2001).<sup>31,32</sup>

Principal Component Analysis is a multivariate statistical technique that is widely used to reduce the number of dimensions of a dataset. So if our data contains a set of  $p$  correlated variables, PCA will transform the set to a smaller set of uncorrelated indices or components, where each component is a linear weighted combination of the initial  $p$  variables.

Given an observed vector of indicator variables  $Y'_{1xp} = [y_1, y_2, \dots, y_p]$  the main goal a PCA will be to create a new set of variables of dimension  $m$  ( $m < p$ ) called principal components. The vector  $Y'$  has mean  $\mu$  and covariance matrix  $\Sigma$  of full rank  $p$ .<sup>34,35</sup>

The principal components ( $z_i$ ) can be written as:

$$z_1 = p_{11}y_1 + p_{12}y_2 + \dots + p_{1p}y_p$$

$$z_2 = p_{21}y_1 + p_{22}y_2 + \dots + p_{2p}y_p$$

$$\vdots$$

$$z_p = p_{p1}y_1 + p_{p2}y_2 + \dots + p_{pp}y_p$$

Where  $p_{ij}$  represents the weight for the  $i$  principal component for the  $j$  variable. In matrix notation  $Z = P'Y$ . To determine the first principal component a vector  $p$  is wanted such that:

$$\begin{aligned} \max \text{var}(z_1) &= \text{var}(p_1'Y) = p_1'\Sigma p_1 \\ \text{st } p_1'p_1 &= 1 \end{aligned}$$

The vector that maximizes this equation (i.e. variance for each principal component) is given by the vector associated with the largest root ( $\lambda_1$ ) of the corresponding eigenvector:

$$|\Sigma - \lambda I| = 0$$

Hence, the first component explains the largest possible amount of variation in the original data, subject to the constraint that the sum of the squared weights  $\sum_{j=1}^p p_{1j}^2$  is equal to one.

As the sum of the eigenvalues equals the number of variables in the initial data set, the proportion of the total variation in the original data set accounted by each principal component is given by  $\lambda/n$ .

The second component is constructed so that is completely uncorrelated with the first component, and explains additional but less variation than the first component, subject to the same constraint. Therefore the covariance between the PC1 and PC2 has to be zero:

$$\text{cov}(z_1, z_2) = p_2'\Sigma p_1 = p_2'p_1\lambda_1 = 0 \Rightarrow p_2'p_1 = 0$$

Given that  $p_2$  is the eigenvector corresponding to the second principal component then:

$$\begin{aligned} \Sigma p_2 &= p_2\lambda_2 \\ \text{var}(z_2) &= \text{var}(p_2'Y) = p_2'\Sigma p_2 = \lambda_2 \end{aligned}$$

Where  $\lambda_2 \leq \lambda_1$ . By Spectral decomposition Theorem, there exists an orthogonal matrix  $P$  such that:

$$P'\Sigma P = \Lambda = \text{diag}[\lambda_i]$$

Where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

This applies for the  $m$  components; therefore, each component captures an additional dimension in the data, while explaining smaller and smaller proportions of the variation of the original

variables. Therefore, the higher the degree of correlation among the original variables in the data, the fewer components required to capture common information.

## 6.2. Statistical Analysis

To estimate the latent variable “socioeconomic status” we chose twenty nine indicators variables included in the National Household Expenditure Survey and performed the principal component analysis on the dataset including the 20.960 sample households. The survey collects data on asset ownership (car, mobile phone, phone), access to utilities and infrastructure (e.g. sewage system and source of water), and housing characteristics (e.g. number of rooms for sleeping and building material). All available variables were included in the analysis.

All of the variables included are categorical variables, except for “number of people per sleeping room”. Categorical variables are not suitable for Principal Component Analysis.<sup>36</sup> Hence, the first step was to recode them generating multiple binary variables for each of the categories.<sup>31,32,36</sup> For this purpose frequencies were ran to construct the indicator variables that will be used, since in some cases frequencies were very low, those categories with low frequencies were grouped together. For example, 93% of the households reported having mostly brick, plastered or paneled wall material. Therefore two categories were created for wall material. All binary variables were coded so that a 1 indicates if the household has a “yes” answer to the question and 0 if it does not have a “no” answer. There were no missing data.

The only continuous variables included was “number of people per sleeping room”, which was constructed by dividing the number of household members by the number of rooms used for sleeping purpose. Following Rutstein, if the number of sleeping rooms is zero, then set the number of sleeping rooms to one since household members have to sleep somewhere.<sup>31</sup>

La Table 1 reports the mean, standard deviation and factor score for the 36 variable included in the analysis. To ensure that there is variation within the population, those variables with low standard deviation were excluded, since that implies that the variable has no variation and therefore would carry low weight in the PCA. If there were variables with zero variance (for eg. all households own a car), then this variable would be the same for every household and would not give information any additional on the relative wealth.<sup>31</sup> If this were the case, such variable would not be useful in the wealth index.

Looking at the first column of Table 1 (mean) it is possible to see that some of the indicator variables have very low frequencies, indicating low ownership, while other variables have very high frequencies, indicating that most of the households own that asset.

The next step was to run the PCA with the 36 selected variables. To compute the index the Principal Component Analysis (*pca*) command in STATA13 was used and only the first component was retained. When constructing wealth or asset index by the principal component analysis method, it is widely accepted that the first component is an adequate measure of welfare.<sup>25,27,31,32</sup>

Given that the variables were not standardized (except for for “number of people per sleeping room”) and they are therefore not expressed in the same units, we ran the analysis using the correlation matrix to ensure that all data have equal weight, this is done by indicating the “*corr*” option. Also we asked the program to provide the mean and standard deviation of each indicator variable.

### 6.3. Results

Principal Component Analysis was performed and the first component was retained, this component accounts for 20.6% of the total variation across the 36 indicators. Scoring factors were divided by their standard deviation of the respective indicator variables so that they would represent the effect a change from categories in the dummy variables.<sup>27,32</sup>

As expected scoring factors for asset ownership have positive sign, indicating that owning that asset increases wealth. For example, a household that live in a house where the block has a sidewalk has a wealth index 0.45 higher that a household that does not. Regarding living conditions, indicators of poor living conditions (households with ground floor or brick loose floor materials, having a septic tank and cesspool, no heating system, etc) have negative scoring factors. In this sense, having ground floor or brick loose floor materials reduces the household wealth index by 0.89.

The most positive effect on the index was given by having a flush toilet (0.86) and water installation in the kitchen (0.80), whereas having the water supply system pipes outside the house but in into yard, plot/compound (0.92) and households with predominant floor material made of ground floor or brick loose (0.89) have the largest negative effect.

Next, the wealth score was estimated by the sum of the indicator variables weighted by the elements of the first eigenvector. Once the wealth measure was obtained, households were classified into quintiles (Poorest, Second, Middle, Fourth, Richest) according to their household wealth index. The



mean score for each quintile was estimated. The means for each indicator variables by wealth quintile are represented in the last five columns of Table 1, the results are in line with the scoring factors.

Those households in the poorest quintiles have higher frequencies in the assets that enter the household wealth index with a negative sign (reduce the score), whereas those households in the higher quintiles show higher frequencies in those indicator variables that contribute positively to the wealth index. This implies that the wealth index estimated with our sample shows internal coherence, this will be further detailed below.

It is important to note that the distribution of the wealth index is not uniform. The difference in the average scores between adjoining quintiles is not even, and this difference decreases as socioeconomic status increases. The last line in Table 1 shows the average wealth index by quintiles. The difference between the poorest and the second poorest is of 3.71. This difference would be reduced by having most of the interior walls revoked or coated (0.64), having source of drinking water from public network (0.47), having the water supply system piped in the interior of the residence (0.94), having a flush toilet (0.86) and having a kitchen with installation of water (0.80). Between the middle and the second poorest is of 1.48, between the fourth and the middle quintile the difference is of 1.07 and between the richest and the fourth is of 0.84. The difference in the wealth index between the richest and the poorest is of 7.11.

Figure 1 shows the distribution of the household wealth index estimated for our data. There is some evidence of truncation at the top. This suggests that the 36 variables included in the analysis were not sufficient to distinguish households among the rich or it could reflect that households are in fact homogenous in terms of SES.

#### *Internal coherence*

Following Filmer and Pritchett (2001) and McKenzie (2003), internal coherence will be examined by analyzing the mean asset ownership differed by socio-economic group. For most of the indicator variables large differences were observed between the poor and the rich.

Let's consider for example "flush toilets", 100% of the households in the richest quintile have flush toilets, while only 57% of the households in the poorest quintile have flush toilets. On the contrary, 37% of the households in the poorest quintile have a water supply system piped outside of residence but into yard, plot/compound, compared to 0% in the highest quintile. Only 11% of the poorest quintile

households own a car, while 66% of the richest quintile does; 16% of the poorest households have sewage system, compared to almost a 100% in the richest quintile.

While the two poorest quintiles use gas cylinder for cooking (89% of the poorest quintile and 81% of the second poorest), the richer quintiles use piped gas (almost a 100% in the richest and 86% of the fourth quintile). Heating system also presents large differences, 62% of the lowest quintile do not have heating system, 97% of the rich have fixed heating appliances.

### *Validation*

The validation of the wealth index was checked with respect to the education variable, which measures the highest educational level of the head of household.<sup>26</sup> In many cases education is used as a proxy for socioeconomic status. Table 2 shows the mean for wealth index as well as the percentage of the households by wealth quintile according to the education level of the head of household.

As can be seen, the economic status varies directly with the educational level of the head of the household. Among those without education, 52% were in the poorest category, while 6% were in the highest wealth quintile. On the contrary, comparing those that attended to the university, 3% correspond to the poorest households and 39% to the richest. In this sense it is possible to conclude that the wealth index predicts socioeconomic inequalities in this sample fairly well.

### *Reliability*

To assess reliability of the household wealth index we used the Cronbach's alpha coefficient. Reliability analysis examines internal consistency among constituent components, i.e. describes the extent to which all the items in a test measure the same concept or construct.<sup>26,37,38</sup> The number of test items, item inter-relatedness and dimensionality affect the value of alpha. According to the literature acceptable values of alpha range from 0.70 to 0.95. A low value of alpha could be due to a low number of questions, poor inter-relatedness between items or heterogeneous constructs. For example if a low alpha is due to poor correlation between items then some should be revised or discarded. If alpha is too high it may suggest that some items are redundant as they are testing the same question but in a different guise. A maximum alpha value of 0.90 has been recommended.<sup>26,38</sup> The Alpha coefficient for our Wealth Index was 0.86, which reflects a good level of reliability of the estimation. Furthermore, we tested whether removing any item would improve the alpha, and it is not necessarily to drop any items.

## 7. Catastrophic health expenditure logistic model

### 7.1. The Logistic regression model<sup>39,40</sup>

The main objective of the logistic regression is to model how the set of predictors (categorical or a mix of categorical and continuous) influence the realization of a random variable. In other words, logistic regression is based on the binomial probability theory in which there are only two plausible outcomes: success or failure. Therefore when our outcome of interest has a binomial distribution then the logistic regression forms a best fitting function using the maximum likelihood method.

This method maximizes the probability of classifying the observed data into the appropriate category given the regression coefficients. There are other methods to solve for the logistic regression, but we will use maximum likelihood method since this is the method STATA uses to obtain the logistic estimation. Therefore only this method will be described in this section.

In order to be able to apply this method it is necessary to construct the likelihood function. This likelihood function expresses the probability of the observed data as a function of the unknown parameters. Solving this function yields the maximum likelihood estimators, which are those parameters that maximize this function.<sup>40</sup>

In the linear regression model, we could express the outcome variable as

$$y = E[Y|x] + \varepsilon$$

Where we assumed that

$$\varepsilon \sim N(0, \sigma_\varepsilon)$$

Then the distribution of  $Y | x$  will be normal with mean  $E[Y|x]$  and a constant variance. In this case we have a dichotomous variable and we can express the outcome

$$y = \pi(x) + \varepsilon$$

Now the error ( $\varepsilon$ ) will assume two possible values:  $\varepsilon = 1 - \pi(x)$  when  $Y=1$ , with probability  $\pi(x)$ ; and,  $\varepsilon = -\pi(x)$  when  $Y=0$ , with probability  $[1 - \pi(x)]$ . Therefore the distribution of the errors will follow a binomial distribution with mean 0 and variance equal to  $\pi(x) [1 - \pi(x)]$ .

Suppose we have  $n$  independent observations  $y_1, y_2, \dots, y_n$  that can be viewed as a realization of a random variable  $Y_i$ . Then we can assume that  $Y_i$  has a binomial distribution:

$$y_i^* \sim B(n_i, \pi_i) \quad (11)$$

Where  $n_i$  is the binomial denominator and  $\pi_i$  is the probability. Then the probability distribution of  $Y_i$  is

$$P[Y_i = y_i] = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (12)$$

With mean and variance given by

$$E(Y_i) = \mu_i = n_i \pi_i$$

$$\text{var}(Y_i) = \sigma_i^2 = n_i \pi_i (1 - \pi_i)$$

Let's consider a set of  $k$  independent variables  $X' = (x_1, x_2, \dots, x_k)$ . We can define the probability of  $Y$  being equal to one given  $X'$  as  $P(Y = 1|X) = \pi(x)$ . Then we can write the logit regression of our multiple logistic model as

$$\pi_i = x_i' \beta \quad (13)$$

As we already mentioned our outcome variable  $Y$  can adopt values 1 or 0, then  $\pi(x)$  can be expressed as follows

$$\pi_i = \frac{\exp\{x_i' \beta\}}{1 + \exp\{x_i' \beta\}} \quad (14)$$

Now we are interested in fitting the logistic regression. The first step will be defining the likelihood function. Equation (10) provides the conditional probability of  $Y$  being equal to 1 given  $x$ ,  $P(Y=1|x)$ . Consequently, the probability of  $Y$  being equal to 0 given  $x$ ,  $P(Y=0|x)$ , will be  $[1 - \pi(x)]$ . For our sample of ordered pairs  $(y_i, x_i)$  it is possible to express the likelihood function for the  $i$ -th pair as

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (15)$$

We assume that the observations are independent, then we can express the likelihood function

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (16)$$

As we mentioned above, the maximum likelihood method estimates  $\beta$  such that the above equation is maximized. Therefore we apply the logarithm to this expression

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (17)$$

Now we can estimate the parameter  $\beta$  that maximized  $L(\beta)$  by differentiating  $L(\beta)$  with respect to  $\beta_0$  and  $\beta_1$  and setting the result equal to zero.

$$\sum_{i=1}^n [y_i - \pi(X_i)] = 0 \quad (18)$$

And

$$\sum_{i=1}^n X_{ij} [y_i - \pi(X_i)] = 0 \quad (19)$$

Where  $j=1, 2, \dots, k$

These expressions are non-linear in  $\beta$ . The estimates obtained are called maximum likelihood estimators and will be denoted as  $\hat{\beta}$ . This leads to the fitted values for the logistic regression

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\pi}(x_i) \quad (20)$$

This means that the sum of the observed values of  $y$  are equal to the sum of the predicted values.

#### **Variance and covariance of the coefficient**

The estimation of the variance and covariance of the coefficients is based on the second partial derivative of the log likelihood function. Below are the estimates of these two derivatives

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (21)$$

And

$$\frac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j} = - \sum_{i=1}^n x_{ij} x_{il} \pi_i (1 - \pi_i) \quad (22)$$

Where  $j, l = 1, 2, \dots, k$ . Now we will define the observed information matrix  $I(\beta)$  as the matrix with dimension  $(k \times 1) \times (k \times 1)$ , containing all the negative terms given in the second order partial derivatives (XXX) and (XXXX). Calculating the inverse of the information matrix we obtain the variances and covariances.

$$VAR(\beta) = I^{-1}(\beta) \quad (23)$$

Then, it is possible to obtain the estimators of  $VAR(\beta)$  by evaluating this at  $\hat{\beta}$ . This estimation will be denoted as  $\widehat{VAR}(\hat{\beta})$  and  $\widehat{Covar}(\hat{\beta}_j, \hat{\beta}_l)$ . The standard error will be given by

$$\widehat{SE}(\hat{\beta}_j) = [\widehat{VAR}(\hat{\beta}_j)]^{1/2} \quad (24)$$

### **The odds ratio**<sup>39</sup>

So far we have talked about probabilities of occurrence of an event. Now we will express those probabilities in terms of odds. Suppose our independent variable (x) is a binary outcome variable. We can rewrite the probability of the event being present given x as

$$Odds_i = \frac{\pi_i}{1-\pi_i} \quad (25)$$

The odds of the outcome being present among individuals with x=1 is defined as

$$Odds(1) = \frac{\pi(1)}{1-\pi(1)} \quad (26)$$

And the odds of the event being present among individuals with x=0 is defined by

$$Odds(0) = \frac{\pi(0)}{1-\pi(0)} \quad (27)$$

Then it is possible to define the odds ratio as

$$OR = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \quad (28)$$

Given that

$$\pi_i = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \quad (29)$$

Then it is possible to substitute in equation (27) and (28)

$$odds(1) = \frac{\frac{e^{\beta_0+\beta_1}}{1-e^{\beta_0+\beta_1}}}{\frac{1}{1+e^{\beta_0+\beta_1}}}$$

$$odds(0) = \frac{e^{\beta_0}}{\frac{1 - e^{\beta_0}}{1 + e^{\beta_0}}}$$

$$OR = \frac{odds(1)}{odds(0)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{(\beta_0 + \beta_1) - \beta_0} = e^{\beta_1} \quad (30)$$

Equation (30) expresses the relationship between the odds ratio and the  $\beta$  coefficients. The odds ratio is the measure of association between the independent variable ( $x$ ) and the dependent variable. It represents how much more likely (or less likely) is for the event to be present among those with  $x=1$  compared to those with  $x=0$ .<sup>39,41</sup>

### ***Testing for significance and goodness of fit***<sup>39,42</sup>

To assess whether the logistic regression coefficients are significant the Wald statistic and likelihood ratio test are used. The Wald statistic is the ratio of the estimated parameters and the corresponding standard error

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Under the null hypothesis that the individual coefficient  $\hat{\beta}_j$  is zero and based on asymptotic theory, this statistic follows a standard normal distribution. If the estimated value of the coefficient is small and its estimated variability is large, then we can not conclude that the estimated coefficient is significantly different from zero.

In the logistic regression the goodness of fit is measured with the deviance. The deviance is the difference of the observed values from the expected values. Therefore, the bigger the difference ("deviance"), then poorer the fit of the model. So, we want this deviance to be as small as possible. Adding more variables to the equation should lead to a smaller deviance, indicating an improvement in the overall fit of the model.

The likelihood ratio test is based on the deviance, where the deviance of the model without the predictor and the deviance with the predictor in the model are compared. The idea behind the test is that removing predictor variables from a model will almost always make the model fit less well. To do so, the log likelihoods of the two models we want to compare are estimated, if this difference is

statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model. The statistic under the hypothesis that the beta's coefficients for the covariates in the model are equal to zero

$$-2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)$$

Where  $L_0$  denotes the likelihood function for the simpler model, and  $L_1$  the likelihood function for the full model. The distribution of this statistic is a chi-square with  $q$  degree-of-freedom, where  $q$  is the number of parameters that are constrained. A significant  $p$ -value provides evidence that the full model performs better than the restricted model.

When assessing goodness of fit of a model we want to test whether the fitted values  $\hat{y}$  differ significantly from the observed, and if the contribution of each pair of  $(y_i, \hat{y}_i)$  is not systematic and small relative to the error structure of the model.<sup>39</sup> There are several measurements of goodness of fit.

**Pseudo  $R^2$ .** There are several  $R^2$  measurements, which basically are variations of the  $R^2$  in linear regression analysis. Among all this alternatives McFadden's  $R^2$  (1973) is one of the most used. It is known as "likelihoodratio index", and compares a model without any predictor to the model including all predictors. It is defined as one minus the ratio of the log likelihood with intercepts only ( $L_0$ ), and the log likelihood with all predictors ( $L_1$ ). If the slope parameters are all 0, then McFadden's  $R^2$  is 0. It will come close (never equal) to 1 if the model is very good.<sup>43</sup>

$$R_{Mc}^2 = 1 - \left(\frac{L_0}{L_1}\right)$$

**Hosmer and Lemeshow Test (2000).** Hosmer and Lemeshow developed a goodness-of-fit test for logistic regression models with binary responses. They proposed to create groups of subjects based on the estimated probabilities and then compare the frequencies of those actually in the each group based on the observed data to the predicted frequencies by the logistic regression model. It is desired that each group contains the same number of observations.<sup>39</sup> This test is computed as a Pearson chi-square statistic from the  $2 \times g$  contingency table of observed and expected frequencies, where  $g$  is the number of groups. The number of groups turns out to be extremely large when including continuous variables. Therefore, usually ten groups are formed (those with estimated probability below 0.1 form one group, those with 0.2 another, and so on up to those with probability 1.0).



$$\hat{C} = \sum_{k=1}^g \frac{(0_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Where  $n'_k$  is the number of observation in the  $i$ th group,  $0_k$  is the number of event outcomes in the  $i$ th group, and  $\bar{\pi}_k$  is the average estimated probability of an event outcome for the  $i$ th group. The Hosmer and Lemeshow statistic is then compared to a chisquare distribution with  $(g - 2)$  degree of freedom.

**Pearson Chi2 Test.** The Pearson residuals are defined to be the standardized difference between the observed frequency and the predicted frequency. They measure the relative deviations between the observed and fitted values

$$\chi^2 = \sum_{j=1}^J \frac{(y_j - m_j \hat{\pi}_j)^2}{v_j}$$

Where

$$v_j = m_j \hat{\pi}_j (1 - \hat{\pi}_j)$$

$m_j$  is the number of subjects with  $X = X_j$  therefore  $\sum m_j = n$ .

## 7.2. The logistic regression applied to catastrophic health expenditure1

In this section we applied the methodology described in section XXX to our analysis. As explained above, the qualitative dependent variable (CHE) in the logistic regression is a dichotomous variable defined as 1 when the household faced catastrophic health payments in the 30 days previous to the survey, and 0 otherwise. Hence it is possible to express the probability of occurrence and non-occurrence of the event (CHE) as

$$P(CHE = 1|X) = F(x, \beta)$$

$$P(CHE = 0|X) = 1 - F(x, \beta)$$

Based on the logistic distribution function, the probability of a household facing catastrophic expenditure is:<sup>1</sup>

$$P = (CHE|X) = P(CHE = 1|X) = F(X'\beta) = e^{X'\beta} (1 + e^{X'\beta})^{-1}$$

where  $X$  is the vector of independent variables and  $\beta$  is the vector of coefficients.

The aim of this study is to determine if the vector of independent variables ( $X$ ) explain whether the household incurs in CHE or not. For this purpose we use a logistic regression model where the set of parameters  $\beta$  reflect the impact of changes in  $x$  on the probability of CHE. Then we can write the probability of a household facing catastrophic health expenditure:

$$P(CHE = 1|X) = \frac{e^{X'\beta}}{1 + e^{X'\beta}} = \Lambda(X'\beta)$$

The odds is the probability of the event of interest to that of the non-event.<sup>41</sup> In our study the event is defined as incurring in catastrophic health expenditure. Therefore the odd will be defined as the proportion of those who experienced catastrophic health expenditure to those who did no given  $X$ .

Under random sampling conditions, a calculated proportion gives us an estimate of the probability of identifying household facing catastrophic payments.<sup>1</sup> Therefore, the 'odds ratio' indicates how often the event happens, relative to how often it does not, under a certain circumstance. The odds ratios (OR) can be written as follows:

$$OR = \frac{P(CHE = 1|X)}{P(CHE = 0|X)} = e^{X'\beta}$$

Applying the logit transformation to equation XXX we obtain

$$\ln\left(\frac{P(CHE = 1|X)}{P(CHE = 0|X)}\right) = X'\beta$$

In this study we have a continuous variable, households OOP payments as a percent of their capacity to pay, which has a discrete realization, is catastrophic when exceeds 40% and is not catastrophic when falls below this threshold. Therefore we can define CHE as

$$CHE = \begin{cases} 1 & \text{if } y^* > 0.4 \\ 0 & \text{otherwise} \end{cases}$$

Then we would model

$$Y^* = X'\beta + E$$

Where the vector of covariates  $X$  contains all the independent variables described in Section 5.2. We assume that  $\epsilon_i$  follows a logistic distribution. The model is estimated by maximum likelihood. As mentioned above, the econometric results also feature the 'odds ratios' that are associated with each

explanatory variable.<sup>17,41</sup> Then the measure of association between our dependent variable (CHE) and the predictors of the model will be given by the odds ratios.

Recall that the OR tells us how much more likely (unlikely) it is for a household to incur in catastrophic health expenditure among those with certain characteristic, considering that the values for the other explanatory variables remain constant.<sup>39</sup> In other words, an odds ratio below 1 for a dependent variable indicates that the factor protects a household from facing catastrophic expenditure, whereas an odds ratio above 1 indicates that the factor is linked to higher probability that the household faces catastrophic expenditure.

When the independent variable is discrete, then the odds ratio represents the marginal effect. It is computed as a discrete probability variation following a change from 0 to 1 for an independent variable, assuming that all other independent variables are constant. For a binary independent variable  $x$ , the expression of the slope is calculated as<sup>1</sup>

$$\frac{\Delta P(CHE = 1|\bar{X})}{\Delta\beta_i} = P(CHE = 1|\bar{X}, \beta_i) - P(CHE = 0|\bar{X}, \beta_i)$$

### 7.3. Statistical analysis

To assess the characteristics of the sample descriptive statistics were calculated (mean, for continuous variables, or proportions, for discrete data) for all households included in the analysis. Logistic regression model was used to identify the determinants of catastrophic expenditure. Z-tests were performed to test the significance of the OR estimated in the logistic regression. Those variables that result non-significant will be dropped, with exception of *coverage* and *Wealth Index*. The reason for dropping variables with non-significant OR is that the purpose of this study is to identify the main determinants of catastrophic health expenditure.

Expenditure can adopt negative values, indicating that the amount sold of certain items exceed the purchase. For the purpose of this analysis those expenses with negative values for any of the spending category were not included. From the total sample 261 observations were deleted.

## 7.4. Results

### 7.4.1. Descriptive analysis

Our sample includes a total of 20.699 households. Data at individual level corresponds to the head of household, data is summarized in Table 3. Overall 65% of the head of households are male, and

62% are married or living together. The average age of the head of households is around 49 years, age increases across wealth groups. The mean age for those in the lowest quintile is 46 years, while for those in the richest quintile is 53 years.

Education also varies across wealth groups. Among those in the poor category, 59% have primary education, this proportion decreases as wealth category increases. The rich are the ones with the highest proportion of head of households that assisted to the university. Employment rate is almost the same for the second, middle and fourth quintile, this are the ones that show the highest employment rates. Those with private insurance have the largest proportion of employed head of households compared with social insurance system and no insurance.

Almost 28% of the households in the study have a senior member (>65 years) this proportion significantly increases with socioeconomic status. In addition, approximately 18% of the households have a senior head of household. Among richest households over 25% of the head of households are over 65 years, compared to almost 15% of the poor households.

The proportion of households with children under five years varies largely across socioeconomic status. Among the poor households almost 55% have at least one child under 5, compared to 30% and 20% in the fourth and richest quintiles respectively. The number of members in the household also decreases significantly as socioeconomic status increases.

#### **7.4.2. Determinants of Catastrophic Health Expenditure**

The average household out-of-pocket health spending in our sample was \$195.41 per month (Table 4). When looking at the different categories of the household health expenditure, the average drug expenditure represents the biggest out-of-pocket payments (\$101.36), representing around 45% of total expenditure, followed by auxiliary services for outpatients (psychology, nursing, speech therapist, etc.) and dental care. Out-of-pocket expenditure varies significantly across socioeconomic status. For those household that are in the richest quintile the OOP payments (\$309.85) is almost 65% higher than that for the poorest quintile (\$108.79).

Looking at Table 4 we can see that, in relative terms, the first quintile spends significantly more on drugs than the richest quintile, in fact drug expenditure decreases as wealth quintiles increase. Out-of-pocket payments in first aid as a percentage of total OOP payments also decrease across wealth quintiles. On the contrary, expenditure on health equipment, dental care, outpatient and inpatient increases across expenditure groups, meaning that these expenditures, as a proportion of total out-of-

pocket health payments, are relatively larger in the richest quintiles than in the lowest income groups. Inpatient services utilization and dental care are fairly constant across income groups.

Out-of-pocket payments as a proportion of the household capacity to pay are highest for the fourth quintile (8.51%), and lowest for the second wealth quintile (1.75%). There are no big differences between this proportion among the poorest, middle and richest quintiles, even when expenditure increases across wealth groups.

The sample includes 28% of households with public sector health coverage, 67% with social insurance system coverage and 5% in the private sector.

When comparing health expenditure for the different types of health insurance, those with private insurance have a health spending that more than threefolds the one of those with no insurance. It is important to remark that payments related to health insurance are not included in the analysis. Out-of-pocket payments for inpatient services are lowest among those without insurance and highest among those with private insurance.

Overall nearly 2.47% of the households incurred catastrophic healthcare expenditure at a capacity to pay threshold of 40%. On the one hand, when stratifying by health coverage, catastrophic health expenditure is significantly higher among those with social insurance system. Almost 3% of the households with social insurance system health insurance incur in catastrophic health expenditure. On the other hand, when stratifying by wealth quintiles differences are not statistically significant, although health expenditures is almost three times higher for the richest households. This difference is explained mainly because of the large capacity to pay the richer households have.

High out-of-pocket payments on health services have had an impact on poverty in Argentina during this period. High proportions of out-of-pockets payments led to financial difficulties for some households, 0.8% of which were pushed into poverty. Differences in impoverishment were significantly different among wealth quintiles but not when comparing between health insurance Fig. 2.

The lowest income group has the highest proportion of households being pushed into poverty (1.2%) due to health payments, Fig. 2. It can be observed in Table 4 that even when the first quintile has catastrophic payments similar to those in the other quintiles, it has the greatest proportion of impoverishment. The fourth wealth quintile was the less affected, only 0.39% were pushed into poverty. The middle and the fifth quintile were also below the overall impoverishment mean.

When looking at impoverishment rate by health insurance it is important to highlight that those with private health insurance are the ones less affected by OOP payments, 0.36% fell below poverty line during the period of analysis Fig. 3. Among those without health insurance 0.89% fell below this threshold, even though this group has the lowest health expenditure.

#### 7.4.3. Logistic regression analysis

Logistic regression was applied to all households in order to explore the determinants of catastrophic expenditure. The binary independent variable is defined as 1 when a household's health expenditure is equal to or above 40% of its capacity to pay and 0 otherwise. The unit of analysis for the regression is the household. In addition to coefficients, marginal effects are presented. In our case the marginal effects are evaluated when a dummy variable is transformed from zero to one.

First we performed a univariate logistic analysis to assess the individual significance of each of the variables included in the analysis. The results are shown in the first column of Table 5. The coefficients are all significant at a 95% confidence with exception with the wealth index. For those variables with more than two categories (education and health insurance type) a Wald test was performed to check its significance (the *test* command after *logistic* was used). Both variables tested significant at 95% confidence.

Secondly we ran the logistic regression including all of the independent variables, including also the wealth index as a continuous variable. Results are shown in table 5 (Model 1). In the complete model several variables tested non-significant: married, education, child and health insurance. We removed them one by one, using backward elimination, starting by those with less significance (higher p-value) and got the final model (Model 2).

Wealth score, which did not showed significance in the univariate analysis was highly significant in the multivariate analysis. This suggests that catastrophic health expenditure is different across wealth quintiles when conditioning for our set of independent variable, i.e. other things being equal. A possible reason for this phenomena is that wealth quintiles are different on some of the social and economic determents included (age, employment status, senior member, senior head of household, poor household and number of members per sleeping room).<sup>44,45</sup> The results suggest that wealth is negatively associated with catastrophic health expenditure. An increase in one unit of the wealth score reduces the chances of incurring in CHE by 0.95 [95%CI 0.92-0.99].

The gender of the household head influences the probability of facing catastrophic payments; specifically, female-headed households are more likely to encounter financial catastrophe than households headed by males. Moreover, those households where the head of household is a male have 0.80 [95%CI 0.66-0.96] chances of incurring in catastrophic health expenditure compared to those where the head of household is a female.

Households that have at least one senior member – aged 65 or more – are 2.14 [95%CI 1.40-3.28] as likely to face catastrophic payments as those with younger head of households. Nevertheless, those households where the head of household is a senior member are less likely to incur in catastrophic payments [OR=0.46, 95%CI 0.29-0.74]. Employment also resulted protective, employed head of households [OR=0.63, 95%CI 0.49-0.80]. Incidence of catastrophic health expenditure is negatively associated with the size of the household.

### ***Goodness of fit***

In this section we will assess the goodness of fit of our model by several tests. The log likelihood chi-square is an omnibus test to see if the model as a whole is statistically significant. It is 2 times the difference between the log likelihood of the current model and the log likelihood of the intercept-only model.<sup>42</sup> The test statistic is 573.85, and that the associated p-value is very low (less than 0.0001). The results show that the addition of the independent variables in the model result in a statistically significant improvement in model fit. Moreover, including these variables to the model reduces the -2 times the log of the likelihood (-2LL) by 573.85, which indicates that this is a very large drop in chi square; ergo we reject the null hypothesis. The effect of at least one of the independent variables likely differs from zero.

Stata also provides McFadden's pseudo R-squared. The statistic suggests the level of improvement over the intercept model offered by the full model, in this case the McFadden's pseudo R2 is of 0.117, which represents a moderate effect.

The Hosmer-Lemeshow statistic suggested a good fitting when splitting the observations into 10 groups. The test statistic was 11.79 with an associated p-value of 0.161, therefore we cannot reject the null hypothesis.

Finally, to determine the model's capacity to discriminate those households that faced catastrophic health expenditure from those that did not, the area below the ROC curve was estimated (Fig. 4). The value of the area is the probability that a household that incurred in CHE had a higher

predicted probability than those households that have CHE=0. For our sample the area was 0.78 (95% C.I. 0.76 to 0.80).

### ***Stratified analysis by health insurance***

One of the secondary objectives of the study was to the study to understand why some households still experience catastrophe and impoverishment due to OOP payments under universal health coverage. In order to test the determinants of CHE varies across type of health insurance a stratified analysis was performed. Although differences in catastrophic health expenditure across the three types of health coverage was significant in the univariate analysis, no statistical differences were observed between coverage type when adjusting for the covariates included in the model (Table 6).

The univariate analysis shows that those with social insurance are 2.35 as likely to experience catastrophe that those with private insurance [95% CI 1.38-4.03]. Those without coverage also showed higher chances of incurring in CHE than households with private insurance, but this difference was not statistically different.

Impoverishment rates also varied according to health coverage. Although different were not statistically significant (Fisher's exact  $p$ -value= 0.175) it is worth mentioning. From the 0.8% of the households that were pushed into poverty, 65% have social insurance and 32% of the households do have health coverage.

## **8. Discussion and conclusions**

This study was based on a representative sample of national population. Out-of-pocket expenditure increases with socioeconomic status measured by the wealth index. When exploring the differences between the associations of catastrophic health expenditure across wealth quintiles no significant differences were found.

Out-of-the pocket expenditure occurs in only 51% of the households, keeping in mind this figures correspond to a 30 day period. There are few households that had health OOP above 40% of their capacity to pay. From the total sample over 97% of the households had OOP that were below this threshold. Albeit poor households show relative low health expenditure, they also show the highest rates of catastrophe together with the rich households. This difference is explained by the low capacity to pay they have, increasing the burden they face from OOP health expenditure. On the contrary



households in the rich quintile have a high CTP and therefore they face a lower burden from OOP payments.

Being employed reduces the odds of CHE. Households that have at least one senior member are more likely to face catastrophic payments. The same is found for those households that have a senior head of household. Age also increases the odds of falling into CHE. On the contrary, being educated and marital status were not significant, therefore were eliminated from the model.

Regarding health coverage, being covered by social insurance system as well as by the public sector is protective of CHE compared to the private sector. For the private sector only 3 variables are significant (age, senior member and senior head of household). For those covered by the social insurance system all variables were significant except for the wealth index and senior head of household. Finally, those households with no insurance do not show association between gender and wealth index.

The three previous studies that analyze the determinants of CHE in Argentina using the NHES have very different findings. First, Peticara (2008) and Cavagnero (2006), both analyze 1996-1997 data and find very different rates of CHE. The former reports a CHE of approximately 8% compared to a 5.5% reported by the later study.<sup>1,2</sup> Knaul et al. (2011) analyzes NHES 2004-2005, in this case 8.4% of the households report CHE. All of these rates are much higher than the ones in our study. One explanation could be differences in measurement tool. Other explanation could be the strengthening of the health system reforms after the crisis in 2001, which could have had an impact in reducing health OOP payments.

The study from Cavagnero et al. study based on the Argentinean Household Expenditure Survey for 1997/1998 had more robust findings than the ones in the present study. The main similarity to this and other studies is the protective effect of the sex of the head of household and employment on CHE, and that having a senior member in the household increases the odds of CHE.<sup>12,14,17</sup>

The analysis from Knaul et al. (2011) found a positive association between the household size, whereas our study found that an increase in the household size was protective against CHE.

Cavagnero et al. (2006) measured SES by income quintiles and found positive association for every quintile compared to the first quintile. This is contrary with the findings of our study where the stratified analysis by wealth quintiles was not significant.<sup>17</sup> Other studies also had similar findings to the ones presented here, where households in lower SES were more likely to experience CHE.<sup>18,20,21</sup>

Out-of-pocket medical expenses for households with private insurance more than twofolds the expenses of households in the Public Sector. It is important to highlight that a limitation of this database is that it only include OOP and there is no information on healthcare service use. This does not allow to adjust for the rate of use of these services, making it hard to address whether households in lower SES have low health expenditure because they have better coverage, or because having a lower CTP discourages the use of healthcare services.

Previous studies in this matter included inpatient and outpatient visits, being both strongly positive associated. As mentioned in the previous paragraph, the NHES does not include healthcare use. Including the expenditure of these two items would have led to endogeneity problems. Therefore they were not included in this analysis.

Impoverishment from OOP health payments was reduced compared to findings from Cavagnero et al. (2006). As mentioned previously he analyzed data from the 1996-1997 and found that 1.7% of households crossed the poverty line after health payments. From this proportion, 52% of those households were in the first quintile and 35% were in the second. Our findings show that impoverishment was reduced by almost half and that the distribution also changed, 31% were from the lowest quintile and 21% corresponded to the second.

One of the biggest limitations of the present study is that the mean predicted probability estimated with the logistic regression is 2.47%, and varies between 0.0001 and 0.3187, which turn out to be very low. The reason for this is that catastrophic health expenditure in our data is a rare event, only 2.46% of the households in the sample experienced CHE. This results systematically biased away from 0 underestimating the true value of  $P(\text{CHE}=1|X)$ . It is recommended that further analysis comparing other approaches would be explored in the future. The literature suggests three main alternative methods<sup>1</sup>:

(i) Bias Correction method proposed by King and Zeng (Stata command: `relogit`. Get it and papers related to it at <http://gking.harvard.edu/relogit> ). This seems to have been very popular with political scientists.<sup>46,47</sup>

(ii) Exact logistic regression (Stata built-in command: `exlogistic`). This only works when N is very small (< 200) and works best when covariates are discrete (preferably dichotomous) and the number of covariates is very small.

---

<sup>1</sup> <http://www3.nd.edu/~rwilliam/stats3/RareEvents.pdf>

(iii) Penalized Maximum Likelihood Estimation proposed by Firth (Stata program: Joseph Coveney's `firthlogit`, available from SSC).<sup>48</sup>

A Monte Carlo simulation comparing the three methods with the MLE estimator agrees that the penalized maximum likelihood estimation (PMLE) provides an unbiased estimation.<sup>49</sup>

Going deeper into these methods exceeds the scope of the present study, and therefore it will be left for future research.

Given the complexity of Argentinean health system it is necessary further research to understand how household in low SES cope with health payments and to what extend health care is covered. Most importantly research aimed to measure to what extend household in low SES tend to reduce the use of health services is necessary.

Further research in this is key since the distributional implications of health systems are a dimension that deserves special attention because it is one of the cornerstones on which these systems are judged.

## 9. References

1. Cavagnero E, Carrin G, Xu K, Aguilar-Rivera AM. Health financing in Argentina: an empirical study of health care expenditure and utilization. *Innovations in Health Financing: Working Paper Series*. 2006(8).
2. Perticara M. *Incidencia de los gastos de bolsillo en salud en siete países latinoamericanos*: CEPAL; 2008.
3. *Objetivos de desarrollo del Milenio. La progresión hacia el derecho a la salud en América Latina y el Caribe*.
4. Kaiser Family Foundation. *Global health facts. Total Expenditure on Health as a percent of gross domestic product, 2005; 2008*.
5. World Health Organization. *The World Health Report 2000. Health systems: improving performance.*; 2000.
6. Censo Nacional Argentino 2010. <http://www.censo2010.indec.gov.ar/>.
7. Ministerio de Salud de la Nación. *Primera Encuesta Nacional de Factores de Riesgo: Informe de resultados. Cobertura y acceso a atención médica*. 2005.
8. Berki SE. A look at catastrophic medical expenses and the poor. *Health Aff (Millwood)*. Winter 1986;5(4):138-145.
9. Joglekar R. Can insurance reduce catastrophic out-of-pocket health expenditure? 2012.
10. Knaul FM, Wong R, Fau - Arreola-Ornelas H, Arreola-Ornelas H, Fau - Mendez O, Mendez O. Household catastrophic health expenditures: a comparative analysis of twelve Latin American and Caribbean Countries. 20110830 DCOM- 20120227 (1606-7916 (Electronic)).
11. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJ. Household catastrophic health expenditure: a multicountry analysis. *Lancet*. Jul 12 2003;362(9378):111-117.
12. Yardim MS, Cilingiroglu N, Yardim N. Catastrophic health expenditure and impoverishment in Turkey. *Health Policy*. Jan 2010;94(1):26-33.
13. Van Doorslaer E, O'Donnell O, Rannan-Eliya RP, et al. Effect of payments for health care on poverty estimates in 11 countries in Asia: an analysis of household survey data. *The lancet*. 2006;368(9544):1357-1364.
14. Van Doorslaer EKA. *Paying Out-of-Pocket for Health Care in Asia: Catastrophic and Poverty Impact*: eSocialSciences;2007.
15. Wagstaff A, Doorslaer E. Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Econ*. 2003;12(11):921-933.
16. Abolhallaje M, Hasani S, Bastani P, Ramezani M, Kazemian M. Determinants of catastrophic health expenditure in Iran. *Iran J Public Health*. 2013;42(Suppl1):155-160.
17. Cavagnero E, Carrin G, Xu K, Aguilar-Rivera AM. *Health financing in Argentina: An empirical study of health care expenditure and utilization*: Fundación Mexicana para la Salud / Instituto Nacional de Salud Pública  
2006.
18. Chuma J, Maina T. Catastrophic health care spending and impoverishment in Kenya. *BMC Health Serv Res*. 2012;12:413.
19. Kim Y, Yang B. Relationship between catastrophic health expenditures and household incomes and expenditure patterns in South Korea. *Health Policy*. May 2011;100(2-3):239-246.
20. Rahman MM, Gilmour S, Saito E, Sultana P, Shibuya K. Health-related financial catastrophe, inequality and chronic illness in Bangladesh. *PLoS One*. 2013;8(2):e56873.
21. Shahrawat R, Rao KD. Insured yet vulnerable: out-of-pocket payments and India's poor. *Health Policy Plan*. May 2012;27(3):213-221.

22. Ukwaja KN, Alobu I, Abimbola S, Hopewell PC. Household catastrophic payments for tuberculosis care in Nigeria: incidence, determinants, and policy implications for universal health coverage. *Infect Dis Poverty*. 2013;2(1):21.
23. Onwujekwe O, Hanson K Fau - Ichoku H, Ichoku H Fau - Uzochukwu B, Uzochukwu B. Financing incidence analysis of household out-of-pocket spending for healthcare: getting more health for money in Nigeria? 20140604 (1099-1751 (Electronic)).
24. Xu K. *Distribution of health payments and catastrophic expenditures Methodology*. . Geneva: WHO/HSS/HSF/HFP Discussion paper 2/05;2005.
25. O'Donnell OA, Wagstaff A. *Analyzing health equity using household survey data: a guide to techniques and their implementation*: World Bank Publications; 2008.
26. Mohanty SK. *Alternative wealth indices and health estimates in India*2009.
27. McKenzie D. Measure inequality with asset indicators, in BREAD working paper No. 042. 2003.
28. Houweling TA, Kunst AE, Mackenbach JP. Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? *International journal for equity in health*. 2003;2(1):8.
29. Bertranou F, Casanova L, Sarabia M. *Dónde, cómo y por qué se redujo la informalidad laboral en Argentina durante el período 2003-2012*: International Labour Organization;2013.
30. Gwatkin DR, Rutstein S, Johnson K, Suliman E, Wagstaff A, Amouzou A. Socio-economic differences in health, nutrition, and population. *Washington, DC: The World Bank*. 2007.
31. Rutstein SO, Johnson K. *The DHS Wealth Index*: DHS Comparative Reports No. 6 2004.
32. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. 20010228 DCOM- 20010503 (0070-3370 (Print)).
33. Okunade AA, Suraratdecha C, Benson DA. Determinants of Thailand household healthcare expenditure: the relevance of permanent resources and other correlates. *Health Econ*. Mar 2010;19(3):365-376.
34. Timm NH. *Applied Multivariate Analysis*: Springer-Verlag New York; 2002.
35. Rencher AC. *Methods of multivariate analysis*. Vol 492: John Wiley & Sons; 2003.
36. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. 20061018 DCOM- 20061214 (0268-1080 (Print)).
37. Caro DH, Cortés D. Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments*. 2012;5:9-33.
38. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education*. 2011;2:53-55.
39. Hosmer Jr DW, Lemeshow S. *Applied logistic regression*: John Wiley & Sons; 2004.
40. Rodríguez G. *Lecture Notes on Generalized Linear Models*. URL: <http://data.princeton.edu/wws509/notes/2007>.
41. Szklo M, Nieto FJ. *Epidemiology: beyond the basics*: Jones & Bartlett Publishers; 2014.
42. Introduction to SAS. Lesson 3 Logistic Regression Diagnostics. *UCLA: Statistical Consulting Group*. from <http://www.ats.ucla.edu/stat/sas/notes2/>.
43. Hu B, Shao J, Palta M. Pseudo-R<sup>2</sup> in logistic regression model. *Statistica Sinica*. 2006;16(3):847.
44. Graham K, Homel R. *Raising the bar*: Routledge; 2012.
45. Katz MH. *Multivariable analysis: a practical guide for clinicians and public health researchers*: Cambridge university press; 2011.
46. King G, Zeng L. Logistic regression in rare events data. *Political analysis*. 2001;9(2):137-163.
47. King G, Zeng L. Explaining rare events in international relations. *International Organization*. 2001;55(03):693-715.

48. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38.
49. Leitgöb H. The Problem of Modeling Rare Events in ML-based Logistic Regression.

## 10. Tables

**Table 1: Summary statistics and scoring factors for indicator variables included in the Wealth Index for present study**

	Summary statistics		Factor Score	Factor Score / SD	Means by quintile					
	Mean	SD			Poorest	Second	Middle	Fourth	Richest	
Most of the interior walls are most revoked or coated	0,900	0,299	0,192	0,641	0,625	0,934	0,966	0,980	0,997	
Predominant walls material	Brick, stone, concrete block	0,934	0,249	0,119	0,478	0,807	0,940	0,945	0,977	1,000
	Natural materials (sun-dried briks, wood, etc)	0,066	0,249	-0,119	-0,478	0,193	0,060	0,055	0,023	0,000
Predominant floor material	Ceramic tile, plastic, wood, rubber, carpet sets	0,797	0,403	0,262	0,651	0,242	0,795	0,956	0,991	0,999
	Cement or brick fixed	0,181	0,385	-0,232	-0,603	0,664	0,197	0,038	0,003	0,000
	Ground floor or brick loose	0,018	0,132	-0,118	-0,892	0,084	0,005	0,001	0,000	0,000
Source of drinking water	Public network	0,962	0,192	0,093	0,486	0,881	0,953	0,982	0,994	1,000
	Other source	0,038	0,192	-0,093	-0,485	0,119	0,047	0,018	0,006	0,000
Water supply system	Piped into residence	0,917	0,276	0,258	0,935	0,591	0,996	0,999	1,000	1,000
	Piped outside of residence but into yard, plot/compound	0,074	0,262	-0,242	-0,923	0,367	0,004	0,001	0,000	0,000
Sewage system	Sewage system	0,644	0,479	0,239	0,499	0,162	0,391	0,748	0,920	0,998
	septic tank and cesspool	0,234	0,423	-0,112	-0,264	0,400	0,493	0,203	0,069	0,002
	Other system	0,123	0,328	-0,205	-0,624	0,438	0,116	0,049	0,011	0,000
Flush toilet	0,908	0,288	0,247	0,856	0,573	0,974	0,995	0,999	1,000	
Household has exclusive use of the bathroom	0,958	0,201	0,072	0,359	0,887	0,953	0,965	0,985	0,999	
Electricity	0,997	0,054	0,034	0,641	0,988	0,999	0,999	1,000	1,000	
Kitchen with installation of water	0,879	0,326	0,260	0,799	0,462	0,947	0,991	0,997	1,000	
Fuel used for cooking	Piped gas	0,496	0,500	0,242	0,484	0,037	0,160	0,427	0,860	0,998
	Gas cylinder	0,469	0,499	-0,223	-0,447	0,892	0,805	0,532	0,116	0,000
	Other sources	0,035	0,183	-0,053	-0,288	0,071	0,036	0,041	0,024	0,002
Heating system	no heating system	0,276	0,447	-0,184	-0,412	0,621	0,402	0,259	0,097	0,002
	fixed appliance	0,471	0,499	0,204	0,409	0,102	0,199	0,380	0,708	0,967
	Mobile heating appliance	0,253	0,435	-0,045	-0,103	0,277	0,399	0,361	0,195	0,031
Air conditioning	0,353	0,478	0,095	0,200	0,134	0,348	0,424	0,352	0,511	
Parking space	0,694	0,461	-0,123	-0,267	0,927	0,776	0,674	0,660	0,432	
Own a cell phone	0,894	0,308	0,040	0,131	0,835	0,901	0,903	0,916	0,916	
Own a land line phone	0,458	0,498	0,167	0,334	0,124	0,309	0,473	0,532	0,855	
The block has improved pavement or gravel	0,707	0,455	0,173	0,380	0,375	0,573	0,756	0,859	0,972	
the house has a sidewalk	0,691	0,462	0,207	0,447	0,252	0,591	0,788	0,845	0,980	
The block has sewage system	0,674	0,469	0,224	0,477	0,236	0,434	0,774	0,931	0,997	
Property ownership	owner	0,648	0,478	0,037	0,077	0,569	0,659	0,642	0,615	0,758
	Renting	0,167	0,373	0,047	0,125	0,087	0,129	0,200	0,247	0,171
	Occupy	0,137	0,344	-0,061	-0,178	0,228	0,163	0,130	0,113	0,052
People sleeping per room	1,904	1,058	-0,140	-0,132	2,566	1,989	1,780	1,706	1,476	
Own a car	0,363	0,481	0,133	0,276	0,110	0,264	0,340	0,443	0,662	
<b>Wealth Index</b>	<b>0,000</b>	<b>2,687</b>			<b>-4,459</b>	<b>-0,742</b>	<b>0,741</b>	<b>1,813</b>	<b>2,655</b>	
Eigen value associated to the first component				7,749		3,716	1,483	1,072	0,842	
Share of the variance associated to the first component				0,206					7,113	
Number of indicator variables used				36						

**Table 2: Validation. Wealth index mean and percentage of households classified into wealth groups by educational level of head of household**

	Wealth Index mean by quintile					Wealth Index mean by quintile				
	Poorest	Second	Middle	Fourth	Richest	Poorest	Second	Middle	Fourth	Richest
No education	-5.095	-0.874	0.850	1.807	2.577	51.47%	19.20%	13.07%	10.67%	5.60%
Primary	-4.590	-0.789	0.727	1.790	2.592	30.88%	22.39%	17.70%	15.74%	13.29%
High School	-4.246	-0.732	0.734	1.825	2.644	16.51%	21.24%	21.65%	21.39%	19.21%
Superior	-3.628	-0.643	0.745	1.802	2.677	7.05%	18.97%	23.46%	23.85%	26.66%
University	-3.344	-0.574	0.783	1.834	2.726	2.75%	9.65%	20.46%	28.09%	39.05%

Note: Other education levels such as special education and kindergarden where omitted, they represento only 0.27% of the total



**Table 3: Descriptive analysis by wealth quintile and health insurance**

	Means by quintiles					P >  z	Means by health insurance			P >  z	All households	
	Poor	Second	Middle	Fourth	Rich		Private	Social insurance	No insurance		Mean	SD
Catartrophic health Expenditure	2.65%	2.24%	2.46%	2.61%	2.67%	0.652	1.34%	3.01%	1.47%	0.000	2.51%	0.157
sex (male)	65.17%	66.79%	64.46%	65.64%	67.15%	0.039	70.45%	65.63%	66.34%	0.001	65.84%	0.474
Married or living together	62.63%	63.39%	60.01%	61.43%	62.35%	0.037	70.36%	61.31%	62.50%	0.000	61.95%	0.486
Age (years)	45.85	47.71	48.60	48.91	53.44	0.000	46.87	51.74	41.99	0.000	48.78	16.193
Education (1)						0.000				0.000		
Primary	58.91%	42.71%	33.76%	29.80%	25.47%		19.82%	36.00%	46.17%		38.23%	0.486
Secondary	31.02%	39.91%	40.66%	39.88%	36.26%		36.34%	36.24%	41.34%		37.59%	0.484
Superior	3.45%	9.30%	11.49%	11.59%	13.12%		13.39%	11.78%	4.50%		9.73%	0.296
University	1.71%	6.01%	12.74%	17.37%	24.45%		29.55%	14.00%	5.47%		12.36%	0.329
other	0.24%	0.34%	0.17%	0.41%	0.19%		0.27%	0.24%	0.32%		0.27%	0.052
Employed (yes)	69.21%	72.42%	71.66%	72.39%	69.02%	0.000	81.61%	67.33%	78.54%	0.000	70.99%	0.454
Elderly (if any member >65 years)	20.85%	24.37%	27.09%	29.50%	39.49%	0.000	20.63%	37.53%	7.38%	0.000	28.03%	0.568
Senior Head of Household	14.83%	16.31%	17.96%	19.50%	25.65%	0.000	12.41%	25.96%	3.10%	0.000	18.71%	0.390
child (if any member < 5 years)	54.74%	38.44%	32.10%	29.70%	20.15%	0.000	32.77%	29.47%	49.86%	0.000	26.95%	0.444
Members in Household	4.03	3.64	3.26	3.19	2.94	0.000	3.19	3.23	3.85	0.000	3.41	1.929
Health insurance (2)						0.000				0.000		
Social Security	43.12%	62.59%	72.84%	76.41%	80.85%						66.87%	0.471
No insurance	54.83%	33.86%	23.12%	17.12%	8.01%						27.69%	0.447

Notes:

(1) Reference no education

(2) Reference private health insurance

**Table 4: Mean Out-of-pocket health paymes by quintiles and health insurance**

	Means by quintile					Statistic*	P >  z	Health insurance			Total	Statistic*	P >  z
	Poorest	Second	Middle	Fourth	Richest			No insurance	Social security	Private			
Health expenditure (OOP)	\$ 108.79	\$ 138.21	\$ 203.47	\$ 223.60	\$ 309.85	42.24	0.0000	\$ 114.91	\$ 221.89	\$ 281.39	\$ 195.41	47.00	0.0000
OOP (% CTP)	4.92%	4.46%	5.18%	4.21%	5.77%	1.12	0.3455	4.03%	5.26%	4.43%	4.89%	2.81	0.0603
Catstrophic health expenditure (% households)	2.66%	2.17%	2.44%	2.53%	2.59%	2.16	0.7060	1.47%	2.94%	1.27%	2.47%	45.48	0.0000
Impovrieshment (% Households)	1.22%	0.83%	0.73%	0.39%	0.69%	16.13	0.0030	0.89%	0.74%	0.36%	0.77%	3.20	0.2020
<b>Expenditures</b>													
<i>Drug</i>	\$ 70.00	\$ 82.66	\$ 100.26	\$ 116.37	\$ 140.51	24.79	0.0000	\$ 55.10	\$ 117.69	\$ 139.03	\$ 101.36	71.76	0.0000
<i>First aid expenditure</i>	\$ 4.55	\$ 5.09	\$ 5.33	\$ 5.45	\$ 5.96	1.00	0.4046	\$ 4.21	\$ 5.71	\$ 6.31	\$ 5.29	4.88	0.0076
<i>Equipment</i>	\$ 5.67	\$ 8.99	\$ 13.88	\$ 18.82	\$ 26.39	41.35	0.0000	\$ 6.50	\$ 17.27	\$ 24.55	\$ 14.65	43.11	0.0000
<i>Outpatient</i>	\$ 8.31	\$ 11.84	\$ 16.31	\$ 15.89	\$ 20.84	21.45	0.0000	\$ 10.75	\$ 15.67	\$ 21.01	\$ 14.54	17.15	0.0000
<i>Dental care</i>	\$ 4.96	\$ 10.18	\$ 21.63	\$ 30.85	\$ 50.50	12.44	0.0000	\$ 9.69	\$ 28.68	\$ 31.82	\$ 23.36	7.17	0.0008
<i>Outpatient, other</i>	\$ 13.03	\$ 15.64	\$ 38.51	\$ 28.22	\$ 46.07	2.98	0.0180	\$ 25.19	\$ 27.78	\$ 37.57	\$ 28.07	0.25	0.7757
<i>Inpatient</i>	\$ 2.25	\$ 3.80	\$ 7.54	\$ 8.00	\$ 19.57	5.26	0.0003	\$ 3.47	\$ 9.10	\$ 21.10	\$ 8.14	4.14	0.0160
Drugs (% OOP)	55.33%	48.66%	43.31%	42.30%	40.35%	34.72	0.0000	50.34%	44.07%	41.77%	45.54%	18.71	0.0000
<i>First aid expenditure (% OOP)</i>	9.50%	8.46%	7.73%	6.45%	5.61%	9.11	0.0000	9.74%	6.88%	7.13%	7.50%	13.00	0.0000
<i>Equipment (% OOP)</i>	7.47%	9.58%	12.15%	14.61%	15.31%	26.84	0.0000	10.22%	12.23%	17.53%	12.10%	15.83	0.0000
<i>Outpatient (% OOP)</i>	13.48%	16.47%	17.80%	15.86%	15.68%	4.88	0.0006	14.35%	16.77%	12.94%	15.90%	7.68	0.0005
<i>Dental care (% OOP)</i>	4.14%	4.70%	5.57%	6.38%	6.18%	4.04	0.0028	4.96%	5.62%	5.11%	5.42%	0.92	0.4000
<i>Outpatient, other (% OOP)</i>	8.82%	10.71%	12.00%	12.28%	14.39%	10.73	0.0000	14.35%	16.77%	12.94%	11.78%	10.56	0.0000
<i>Inpatient (% OOP)</i>	1.25%	1.43%	1.44%	2.11%	2.47%	4.86	0.0006	0.98%	1.92%	3.08%	1.77%	10.12	0.0000

\* Anova or Chi2, depending on wheather the avriable was binary or continuous

**Table 5: Logistic regression for catastrophic health expenditure**

	Univariate analysis		Modelo 1		Modelo 2	
	Coefficient (SD)	P> z	OR (SD)	P> z	OR (SD)	P> z
Male	0.520 █ (0.046)	0.000	0.674 █ (0.078)	0.001	0.740 █ (0.072)	0.002
Married	0.493 █ (0.044)	0.000	1.133 █ (0.153)	0.353	1.045 █ (0.005)	0.000
Age	1.056 █ (0.003)	0.000	1.047 █ (0.005)	0.000		
Education						
<i>Elementary</i>	0.940 █ (0.262)	0.823	1.227 █ (0.356)	0.480		
<i>High school</i>	0.529 █ (0.15)	0.025	1.286 █ (0.391)	0.408		
<i>Superior</i>	0.513 █ (0.163)	0.035	1.154 █ (0.392)	0.674		
<i>University</i>	0.402 █ (0.127)	0.004	1.094 █ (0.375)	0.794		
<i>Special</i>	0.974 █ (0.749)	0.972	2.078 █ (1.638)	0.354		
Employed	0.352 █ (0.027)	0.000	0.724 █ (0.064)	0.000	0.717 █ (0.061)	0.000
Senior	2.474 (0.134)	0.000	1.381 (0.185)	0.016	1.387 (0.168)	0.007
Head Of household senior	4.905 (0.437)	0.000	0.598 (0.138)	0.026	0.602 (0.129)	0.018
Child	0.335 (0.047)	0.000	1.188 (0.209)	0.326		
Number of memebers	0.698 (0.022)	0.000	0.765 (0.031)	0.000	0.791 (0.025)	0.000
Poor	0.347 (0.072)	0.000	0.232 (0.05)	0.000	0.234 (0.049)	0.000
Health Insurance						
<i>Social security</i>	2.282 (0.604)	0.002	1.547 (0.418)	0.106		
<i>Public sector</i>	1.101 (0.311)	0.733	1.744 (0.516)	0.060		
Wealth index	1.014 (0.017)	0.408	0.935 (0.019)	0.001	0.927 (0.016)	0.000
Constant	0.014 (0.017)	0.408	0.004 (0.002)	0.000	0.008 (0.002)	0.000

**Categorical variables Tests:**

Education	chi2= 43.510	chi2= 1.970
	P > chi2= 0.000	P > chi2= 0.853
Health Insurance	chi2= 52.320	chi2= 3.540
	P > chi2= 0.000	P > chi2= 0.170

**Table 6: Stratified Logistic regression for catastrophic health expenditure**

	Private			Social Security			Public		
	OR	P> z	95% CI	OR	P> z	95% CI	OR	P> z	95% CI
Male	2.538	0.188	[0.634;10.16]	0.758	0.011	[0.612;0.937]	0.752	0.230	[0.472;1.198]
Age	1.112	0.006	[1.031;1.2]	1.035	0.000	[1.023;1.048]	1.054	0.000	[1.033;1.075]
Employed	0.453	0.311	[0.098;2.099]	0.731	0.042	[0.54;0.988]	0.486	0.004	[0.298;0.793]
Senior	15.257	0.002	[2.721;85.543]	1.895	0.023	[1.091;3.291]	2.124	0.070	[0.94;4.799]
Head Of household senior	0.065	0.011	[0.008;0.537]	0.620	0.114	[0.343;1.121]	0.217	0.011	[0.067;0.708]
Number of memebers	1.069	0.703	[0.758;1.507]	0.783	0.000	[0.725;0.847]	0.821	0.001	[0.727;0.926]
wealths Index	0.904	0.445	[0.698;1.171]	0.964	0.106	[0.923;1.008]	0.944	0.101	[0.881;1.011]
Constant	0.000	0.000	[0;0.002]	0.010	0.000	[0.005;0.021]	0.005	0.000	[0.001;0.016]

## 11. Figures

Figure 1: distribution of the household wealth index

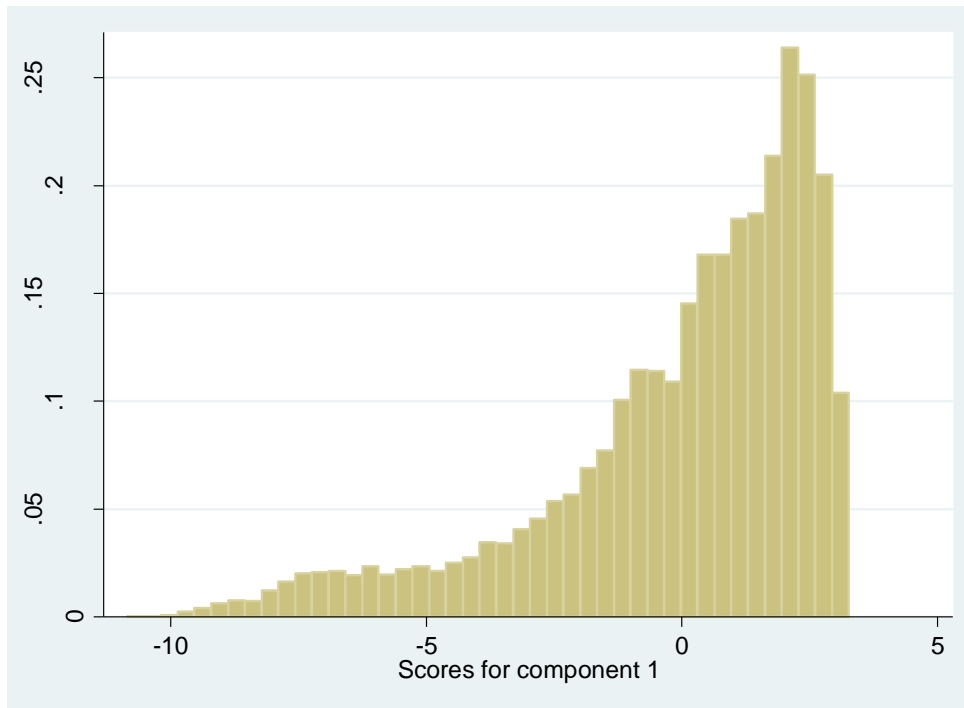


Figure 2: Impoverishment rate after out-of-pocket payment by wealth quintiles

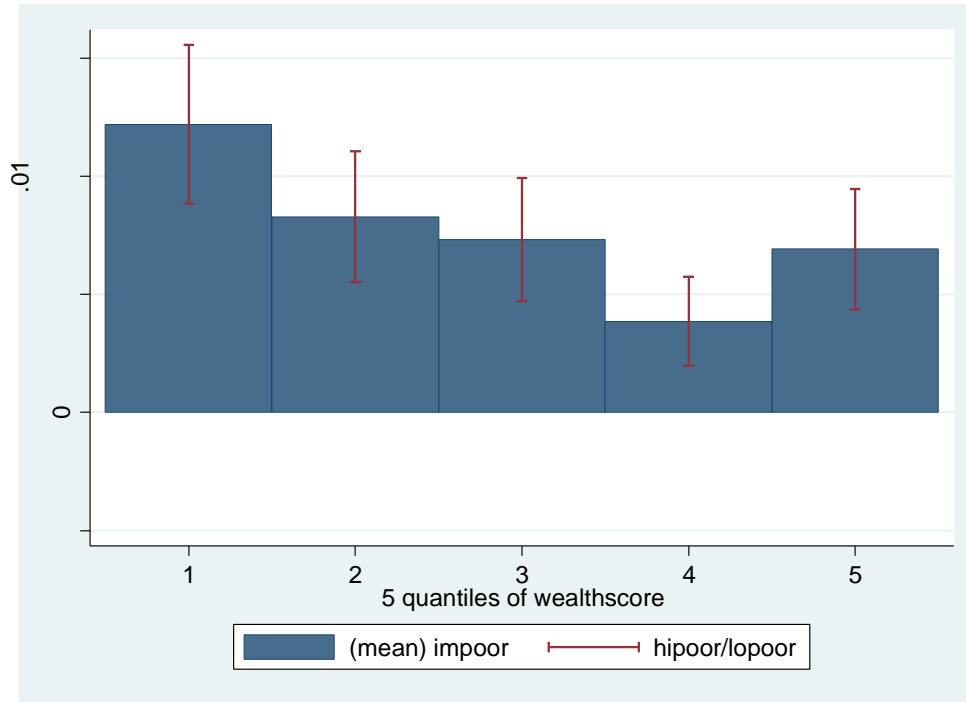


Figure 3: Impoverishment rate after out-of-pocket payment by health insurance

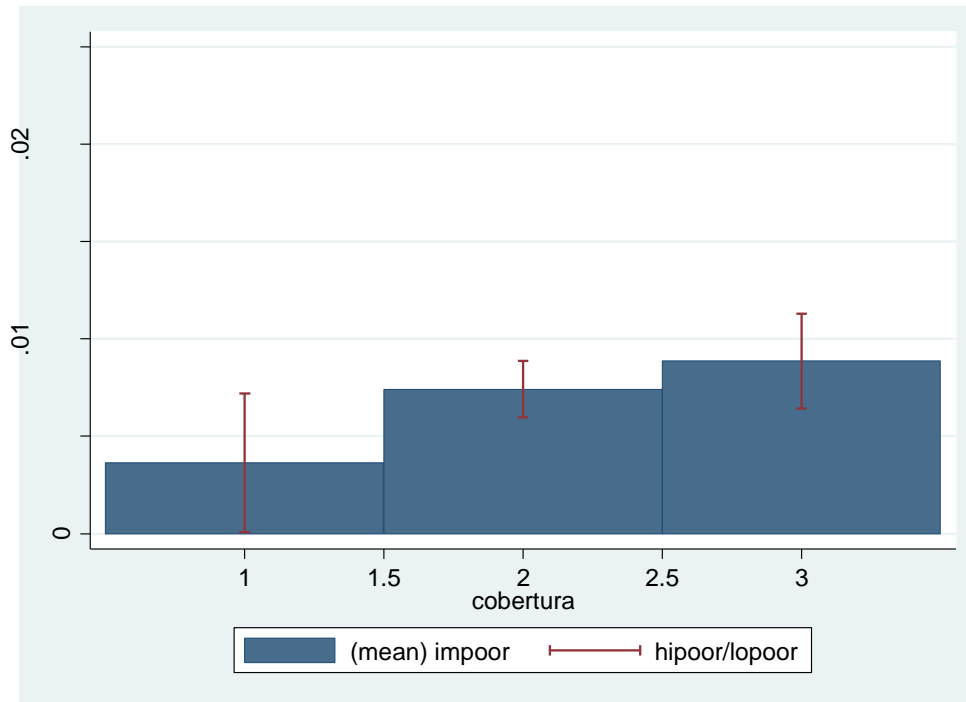


Figure 4: Roc Curve – Model calibration

